

APPENDIX F—STANDARD SETTING REPORT



2008

New England Common Assessment Program

Grade 11 Standard-Setting Report

**January 9 & 10, 2008
Portsmouth, New Hampshire**

Reading/Mathematics -	61
APPENDIX G: NECAP Grade 11 Final Writing Rubrics.....	65
APPENDIX H: NECAP Standard Setting Grade 11 Rating Forms - Writing Rounds 1 and 2 -.....	71
APPENDIX J: NECAP Standard Setting Evaluation Summaries	75
APPENDIX K: NECAP Standard Setting Panelists.....	87

Standard-Setting Process

The standard-setting meeting, to establish cut scores for the grade 11 *NECAP* in reading, writing, and mathematics, was held on Wednesday and Thursday, January 9 & 10. Each content area panel consisted of 16 or 17 participants.

A modified version of the Bookmark standard-setting method was implemented for all grades in mathematics and reading. A modified version of the Body of Work method was used for writing. An overview of the methods is described below.

To help ensure consistency of procedures between panels, each panel was led through the standard-setting process by a trained facilitator from Measured Progress.

OVERVIEW OF PROCESS

This section of the report provides an overview of the standard-setting process as implemented for *NECAP*. The process was divided into three stages, each with a number of constituent tasks.

1. Tasks completed prior to the standard-setting meeting
 - Creation of achievement level descriptions
 - Collection and analysis of existing performance data
 - Calculation of starting cut-points for writing
 - Preparation of materials for panelists
 - Preparation of presentation materials
 - Preparation of *Instructions for Facilitators Documents*
 - Preparation of systems and materials for analysis during the meeting
 - Selection of panelists
2. Tasks completed during the standard-setting meeting
 - Orientation
 - Reading and mathematics:
 - Review of assessment materials
 - Completion of item map
 - Review of achievement level descriptions (ALDs) and definition of borderline students
 - Round 1 judgments—mathematics
 - Tabulation of Round 1 results—mathematics
 - Round 2 judgments—mathematics
 - Tabulation of Round 2 results—mathematics
 - Round 3 judgments—mathematics
 - Round 1 judgments—Reading
 - Tabulation of Round 1 results—Reading
 - Round 2 judgments—Reading

- Writing
 - Discussion of writing scoring rubrics and anchor papers
 - Review of general achievement level descriptions
 - Review and discussion of starting cut-points
 - Writing to the common prompt
 - Round 1 judgments—common prompt
 - Tabulation of Round 1 results
 - Round 2 judgments—common prompt
 - Repeat Rounds 1 and 2 for each matrix prompt
 - Round 3 judgments
 - Evaluation
- 3. Tasks completed after the standard-setting meeting
 - Analysis and review of panelists' feedback
 - Preparation of recommended cut scores
 - Preparation of standard-setting report

1. TASKS COMPLETED PRIOR TO THE STANDARD-SETTING MEETING

1.1 Creation of Achievement Level Descriptions (ALDs)

The ALDs presented to panelists provided the official description of the set of knowledge, skills, and abilities that students are expected to display in order to be classified into each achievement level. The descriptions are provided as Appendix A of this document.

1.2 Collection and Analysis of Existing Performance Data

Prior to standard setting, a variety of data was gathered and examined for possible use in establishing starting cut-points for reading and mathematics. (A different method was used for writing; see the section that follows.) These data sources included:

- Teacher judgment data, collected from the students' grade 10 teachers prior to the administration of the assessment in the fall;
- Performance of students on the reading and mathematics tests in grades 6 through 8 and
- Performance on high school-level tests given in prior years

Teacher Judgment Data. In the spring of 2007, teachers of grade 10 students were asked to review the descriptions of the four achievement levels and to rate their students based on classroom performance. A web site was created for teachers to enter their ratings. While this method of collecting the data is not ideal, it was not feasible to record the ratings directly on the students' test booklets, as was done in 2006 for grades 3 through 8, primarily because grade 11 teachers would not have been familiar enough with the students to rate them accurately. Because of this data collection method and because of difficulties encountered in matching teacher judgment data to students' test scores, data were only obtained for approximately 10% of the students tested. This amount of data was considered too sparse for starting cut-points, and so was not used.

Existing Test Data. Two categories of existing test data were examined: 1) fall 2007 scores in grades 6 through 8 and 2) historical performance on other high school-level tests (for example, NAEP).

For reading, starting cut-points were calculated from the existing test data as follows: the pattern of performance on the fall 2007 *NECAP* reading tests in grades 6, 7, and 8, was determined (specifically, the percentage of students in each achievement level category). Predicted grade 11 scores were then calculated by extrapolation. The resulting cuts were found to be in line with other high school-level testing data and to represent reasonable starting points. Therefore, they were adopted as starting cuts for standard setting. The starting cuts were presented to panelists as placements in the ordered item booklet (see below for complete details), and panelists were asked to either validate the placements or recommend modifications.

For mathematics, potential starting cuts were calculated in the same way as for reading, but were not used for standard setting. The purposes of using starting cuts are to streamline and simplify the standard-setting process and to make use of any other relevant sources of available information. However, the grade 11 mathematics test was quite difficult for the students, and the extrapolated starting placements for the lower two cuts appeared very early in the ordered item booklet (specifically, between ordered items 1 and 2 and between ordered items 6 and 7). This anomaly

suggested that differences between the grade 11 mathematics test and the previously existing data rendered the use of those data, and the resulting cuts, inappropriate. In addition, it was feared that the use of such low starting cuts would complicate the process for the panelists and possibly impact the validity of the results negatively. For these reasons, a standard-setting, rather than a standards-validation, approach was adopted for mathematics.

1.3 Establishing Starting Cut-points for Writing

Reading consultants from each of the three state departments met to discuss starting cut-points for standard setting. It was determined that the starting cut-points would be established based on the scoring rubric and its relationship to the achievement level definitions. The states set the following score ranges as best representing the language of the achievement level definitions and these were used as starting cut-points:

Achievement Level	Raw Score Cuts
Proficient/Proficient with Distinction	9/10
Partially Proficient/Proficient	6/7
Substantially Below Proficient/Partially Proficient	3/4

1.4 Preparation of Materials for Panelists

The following materials were assembled for presentation to the panelists at the standard-setting meeting:

- Meeting agenda
- Confidentiality agreement
- ALDs
- Assessment booklet
- Answer key/scoring rubrics
- Ordered item booklet (reading and mathematics)
- Item maps (reading and mathematics)
- Bodies of Work (writing)
- Rating forms
- Evaluation form

1.5 Preparation of Presentation Materials

The PowerPoint presentation used in the opening session was prepared prior to the meeting. A copy of the PowerPoint slides is included as Appendix B of this document

1.6 Preparation of Instructions for Facilitators Documents

For each content area, a document was created for the group facilitator to refer to while working through the process. The version for reading is included as Appendix C, the version for mathematics as Appendix D, and the version for writing as Appendix E.

1.7 Preparation of Systems and Materials for Analysis During the Meeting

The computational programming to carry out all analyses during the standard-setting meeting was completed and thoroughly tested prior to the standard-setting meeting.

1.8 Selection of Panelists

Panelists were selected prior to the standard-setting meeting by the client states. The goal was to recruit 18 teachers for each panel, six from each state. Because *NECAP* is administered in the fall and is designed to measure grade level expectations for the end of the previous grade, it was decided that four of the six from each state should be from grade 11 and two should be from grade 10. These criteria were followed as closely as possible in recruiting and selecting the panelists. The majority of the panelists were general education teachers, but some special education and ESL teachers were recruited as well.

The actual number of panelists who participated was 49, 16 each in the reading and writing groups, and 17 in the mathematics group. Of these, 18 were from New Hampshire, 17 were from Vermont, and 14 were from Rhode Island. Panelists from each state were distributed fairly uniformly across the different panels. (List of panelists is included as Appendix K.)

2. TASKS COMPLETED DURING THE STANDARD-SETTING MEETING

2.1 Orientation

The standard-setting meeting began with a general orientation session that was attended by all panelists. The purpose of the orientation was to provide background information, an introduction to the issues of standard setting, and a brief overview of the activities that would occur during the standard-setting meeting. Once the general orientation was complete, the writing panelists reconvened into their breakout room, where they received training specific to the Body of Work method and began the rating process. The reading and mathematics groups remained together and were given an overview of the bookmark process, after which they reconvened in their breakout rooms.

Because the process followed by writing was somewhat different than that followed by reading and mathematics, the remainder of this section of the report is presented by content area. In addition, there are some differences between the processes followed by the reading and mathematics groups, so some subsections are further broken out by the two areas.

2.2 Mathematics and Reading

2.2.1 Review of Assessment Materials

Once the reading and mathematics panels convened in their breakout rooms, the first step was to take the test for their content area. The purpose of this step was to make sure the panelists were thoroughly familiar with what the assessment asks of students. Once panelists completed the test an answer key was distributed. At this point, panelists were encouraged to discuss any issues that came to mind regarding items or scoring.

2.2.2 Completion of Item Map

The purpose of the next step was to ensure that panelists became very familiar with the ordered item booklet and understood the relationships among the ordered items. The ordered item booklet contained one item (or item-score category) per page, ordered from the easiest to the most difficult. The ordered item booklet was created by sorting items by their IRT-based difficulty values (b corresponding to $RP_{0.67}$ was used). A three-parameter logistic IRT model was used for the dichotomous items and the graded response IRT model was used for the polytomous items. The group facilitators explained to the panelists that each open-response item would appear multiple times in the ordered item booklet, once for each possible score point.

The item map listed the items in the same order they were presented in the ordered item booklet and had spaces for the panelists to write in the knowledge, skills, and abilities required to answer correctly (or earn a particular score point). There was also a space for the panelists to write in why they felt the current ordered item was more difficult than the previous one.

Because starting cuts were used for reading, and because the item mapping process can be very time-consuming, the task was narrowed for reading panelists by instructing them to start approximately five ordered items prior to each starting cut-point and stop approximately five ordered items after the cut. The range of plus or minus five ordered items was a guideline only, and panelists were free to expand that range as appropriate. For the mathematics panel, where no starting cuts were used, it was necessary for panelists to complete the item map for the full item set.

Each panelist stepped through the ordered item booklet, item by item, considering the knowledge, skills, and abilities students needed to complete each one. They recorded this information onto the item map along with reasons why an item was more difficult than the previous one. After they were finished working individually, panelists had an opportunity to discuss the item map as a group and make necessary additions or adjustments.

2.2.3 Review of ALDs and Definition of Borderline Students

Next, panelists reviewed the ALDs. This important step of the process was designed to ensure that panelists thoroughly understood the needed knowledge, skills, and abilities to be classified as *Partially Proficient*, *Proficient*, and *Proficient with Distinction*. Panelists began individually then discussed the descriptions as a group, clarifying each level. Afterwards, panelists developed consensus definitions of borderline students, i.e., students who are “just able enough” to be categorized into an achievement level. Bulleted lists of characteristics for each level were generated based on the whole group discussion and posted in the room for reference throughout the bookmark process.

2.2.4 Round 1 Judgments—Mathematics

In the first round, panelists worked individually with the ALDs, the item map they completed earlier, and the ordered item booklet. Beginning with the first ordered item, and considering the skills and abilities needed to complete it, they asked themselves the question, “Would at least 2 out of 3 students performing at the borderline of *Partially Proficient* answer this question correctly (or earn this score point)?” Panelists considered each ordered item in turn, asking themselves the same question until their answer changed from “yes” (or predominantly “yes”) to “no” (or predominantly “no”). A bookmark was placed there. Panelists then repeated the process for the other two cuts and used the provided rating form to record his/her ratings for each cut (see Appendix F).

2.2.5 Tabulation of Round 1 Results—Mathematics

After the Round 1 ratings were complete, Measured Progress staff calculated the average cut-points for the room based on Round 1 bookmark placements. This information was shared with the group to assist them in Round 2.

2.2.6 Round 2 Judgments—Mathematics

The purpose of Round 2 was for panelists to discuss their Round 1 placements and revise their ratings, if necessary. Panelists shared their individual rationales for their bookmark placements in terms of the necessary knowledge and skills for each classification. Panelists were asked to pay particular attention to how their individual ratings compared to those of the others and get a sense for

whether they were unusually stringent or lenient within the group. Room average cut-points were to be considered as well.

Although the panelists worked as a group, the facilitators made sure it was understood that they should set the bookmark according to their *individual* best judgments, and that they need *not* come to consensus. They were encouraged to listen to the points made by their colleagues but not feel compelled to change their bookmark placements.

Finally, panelists were given the opportunity to revise their Round 1 ratings on the rating form.

2.2.7 Tabulation of Round 2 Results—Mathematics

When Round 2 ratings were complete, Measured Progress staff calculated the average cut-points for the room and associated *impact* data. Impact data gave the percentage of students across the three states that would fall into each achievement level category according to the cut-points. This information was shared with the group to assist them in Round 3.

2.2.8 Round 3 Judgments—Mathematics

The purpose of Round 3 was to give panelists a final opportunity to discuss and, if necessary, modify their bookmark placements. Panelists were asked to consider all Round 2 results and the input of their colleagues. Once again, facilitators made sure panelists understood they were providing individual bookmark placements and not coming to consensus.

After the group discussions, panelists once again recorded bookmark placements on the rating form.

2.2.9 Round 1 Judgments—Reading

For reading, starting cut-points were provided to panelists. This effectively took the place of the final, individual, round of ratings as implemented for mathematics. Reading panelists worked as a group in their first round, evaluating and (if necessary) revising the starting cut-points. Using the ALDs, the item map they completed in the previous step, and the ordered item booklet, they began with the ordered item approximately five items before the *Partially Proficient* starting cut-point, and considering the skills and abilities needed to complete it, asked themselves the question, “Would at least 2 out of 3 students performing at the borderline of *Partially Proficient* answer this question correctly (or earn this score point)?” Panelists considered each ordered item in turn, asking themselves the same question until their answer changed from “yes” (or predominantly “yes”) to “no” (or predominantly “no”). A bookmark was placed there. Panelists then repeated the process for the other two cuts and used the provided rating form to record his/her ratings for each cut (see Appendix F).

Although the panelists worked as a group, the facilitators made sure it was understood that they should set the bookmark according to their *individual* best judgments, and that they need *not* come to consensus. They were encouraged to listen to the points made by their colleagues but not feel compelled to change their bookmark placements.

2.2.10 Tabulation of Round 1 Results—Reading

When Round 1 ratings were complete, Measured Progress staff calculated the average cut-points for the room and associated *impact* data. Impact data gave the percentage of students across the three states that would fall into each achievement level category according to the cut-points. This information was shared with the group to assist them in Round 2.

2.2.11 Round 2 Judgments—Reading

The purpose of Round 2 was to give panelists an opportunity to discuss and, if necessary, modify their Round 1 bookmark placements. Panelists shared their individual rationales for their bookmark placements in terms of the necessary knowledge and skills for each classification. Panelists were asked to pay particular attention to how their individual ratings compared to those of the others and get a sense for whether they were unusually stringent or lenient within the group. Room average cut-points were to be considered as well.

Finally, panelists were given the opportunity to revise their Round 1 ratings on the rating form.

2.3 Writing

2.3.1 Discussion of Writing Scoring Rubrics and Anchor Papers

The writing panelists began by reviewing the five writing scoring rubrics: Response to Literary or Informational Text, Reflective Essay, Persuasive Essay, Report, and Procedure (see Appendix G). Particular attention was paid to the rubric for Response to Informational Text, since that was the genre for the common prompt.

2.3.2 Review of General ALDs

Next, panelists reviewed the general ALDs. This important step of the process was designed to ensure that panelists thoroughly understood the needed knowledge, skills, and abilities to be classified as *Partially Proficient*, *Proficient*, and *Proficient with Distinction*. Panelists began individually and afterwards discussed the descriptions as a group, clarifying each level. Consensus definitions of students at each level were made into bulleted lists that were kept posted in the room for reference throughout the process.

2.3.3 Review and Discussion of Starting Cut-Points

Next, the facilitator described the process used to determine the starting cut-points, after which panelists discussed then and provided feedback or proposed alternatives.

2.3.4 Writing to the Common Prompt

Next, panelists wrote to the common prompt. The purpose of this step was to make sure they were thoroughly familiar with what the prompt asked students to do.

2.3.5 Round 1 Judgments—Common Prompt

The panelists were given a set of 16 student papers (responses to the common prompt) to for making their ratings. The papers were presented in order (from lowest scoring to highest), but the scores themselves were not revealed during the Round. Working individually, panelists reviewed each paper for the skills and abilities demonstrated and their relationship to the ALDs. Panelists categorized each paper into one of the four levels, recording them on the rating sheet. (A sample of the rating sheets used for writing Rounds 1 and 2 is included as Appendix H.)

2.3.6 Tabulation of Round 1 Results

When Round 1 ratings were complete, Measured Progress staff calculated the average cut-points for the room. This information was shared with the group to assist them in Round 2.

2.3.7 Round 2 Judgments—Common Prompt

The purpose of Round 2 was for the panelists to discuss and, if necessary, revise their Round 1 ratings. They were provided with the room average cut-points from Round 1 and the scores awarded to each paper. Prior to beginning the Round 2 discussions, using a show of hands, the room facilitator indicated on chart paper how many panelists assigned each paper to the achievement levels. The facilitator also indicated on the chart paper how each paper would be categorized based on the Round 1 room average cut-points.

Beginning with the first paper for which there was disagreement on categorization, panelists shared their individual rationales for categorization. The panelists were asked to pay particular attention to how their ratings compared to those of the others and get a sense for whether they were unusually stringent or lenient within the group.

After the discussion, panelists were given the opportunity to revise their Round 1 ratings in the Round 2 column on the rating form. Facilitators reminded panelists that their best individual judgment was wanted, and that no one should feel compelled to change their ratings.

2.3.8 Repeat Rounds 1 and 2 for Each Matrix Prompt

After completing Rounds 1 and 2 for the common prompt, the panel followed virtually the same process for each of the five matrix prompts one by one, completing both rounds of ratings for one before proceeding to the next. The process differed from that used for the common prompt in two ways: first, panelists were asked to rate a set of 11 papers (one per score point, from 2 through 12) rather than the 16 used for the common prompt; and, second, the panelists knew the score awarded to each paper prior to doing their Round 1 ratings.

2.3.9 Round 3 Judgments

After Rounds 1 and 2 were complete for the common and all five matrix prompts, panelists were given one last opportunity to discuss the placement of the cuts or any remaining issues. Then they were asked to recommend a single set of raw score cut-points to be used for all prompts on the Round 3 rating form (See Appendix I).

2.4 Evaluation

As the last step in the standard-setting process, panelists in all three groups anonymously completed an evaluation form. The results of the evaluations are presented as Appendix J.

3. TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING

Upon conclusion of the standard-setting meeting, several important tasks were completed. These tasks centered on reviewing the standard-setting meeting and addressing anomalies that may have occurred in the process or in the results.

3.1 Analysis and Review of Panelists' Feedback

Upon completion of the evaluation forms, panelists' responses were reviewed. This review did not reveal any anomalies in the standard-setting process or indicate any reason that a particular panelist's data should not be included when the final cut-points were calculated. It appeared that all panelists understood the rating task and attended to it appropriately.

3.2 Preparation of Recommended Cut Scores

After the standard setting was completed, the cut-points on the ordered item scale and on the theta (θ) scale were calculated for mathematics based on the panelists' Round 3 ratings, and for reading based on the Round 2 ratings. In addition, the percentages of students who would be classified into each achievement level were determined. These results are presented in Tables 1 and 2 below in the columns labeled "Standard Setting Recommended Cuts." Table 1 also shows the corresponding information for the starting cuts used for reading.

Table 1: Summary of *NECAP* Standard-Setting Results—Reading

Achievement Level	Starting Cut Points		Standard Setting Recommended Cuts		
	Raw Score Range	% in Category	Raw Score Range	Theta Cut	% in Category
Proficient with Distinction	40-52	13.7	39-52	1.0038	17.4
Proficient	29-39	47.8	28-38	-0.3099	47.8
Partially Proficient	19-28	25.9	19-27	-1.2071	22.3
Substantially Below Proficient	0-18	12.5	0-18		12.5

Table 2: Summary of *NECAP* Standard-Setting Results—Mathematics

Achievement Level	Standard Setting Recommended Cuts		
	Raw Score Range	Theta Cut	% in Category
Proficient with Distinction	53-64	2.0586	1.5
Proficient	29-52	0.6190	24.5
Partially Proficient	18-28	-0.1169	27.5
Substantially Below Proficient	0-17		46.5

For writing, the final recommended cuts, based on the panelists' Round 3 ratings, are shown in Table 3. The table also shows the corresponding percentages in each category. Note that the cuts recommended by the panelists were the same as those recommended by the content experts and used as starting cuts.

**Table 3: Summary of *NECAP* Standard-Setting Results—
Writing**

Achievement Level	Standard Setting Recommended Cuts	
	Raw Score Range	% in Category
Proficient with Distinction	10-12	3.3
Proficient	7-9	32.2
Partially Proficient	4-6	48.3
Substantially Below Proficient	0-3	16.1

3.3 Preparation of Standard-Setting Report

Following final compilation of standard-setting results, Measured Progress prepared this report, which documents the procedures and results of the 2008 standard-setting meeting in order to establish performance standards for the Grade 11 New England Common Assessment Program (*NECAP*) in reading, mathematics and writing.

APPENDICES

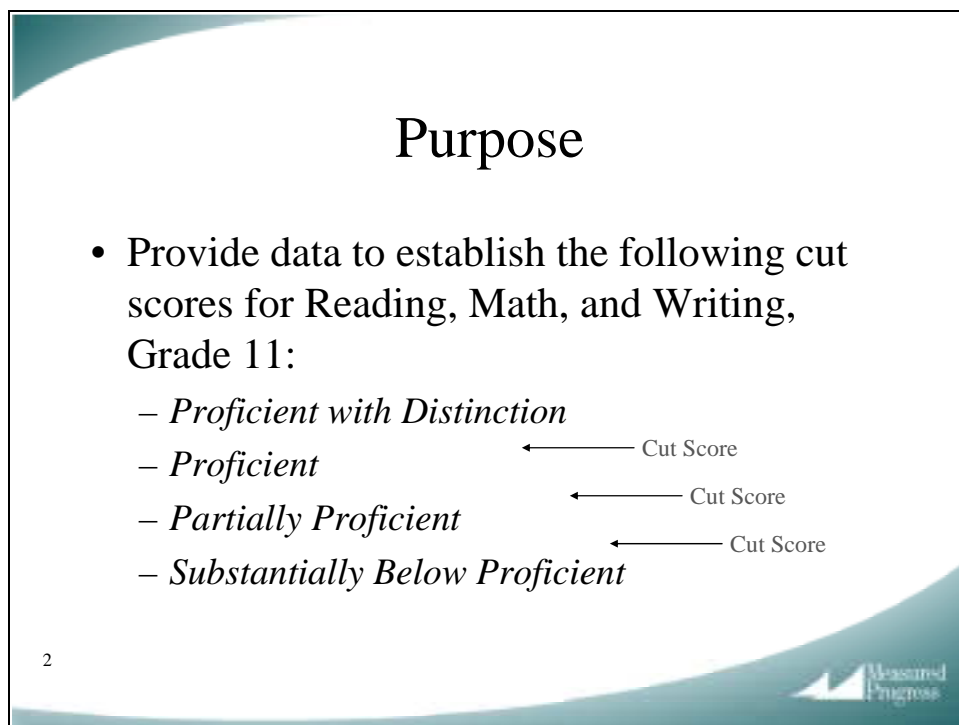
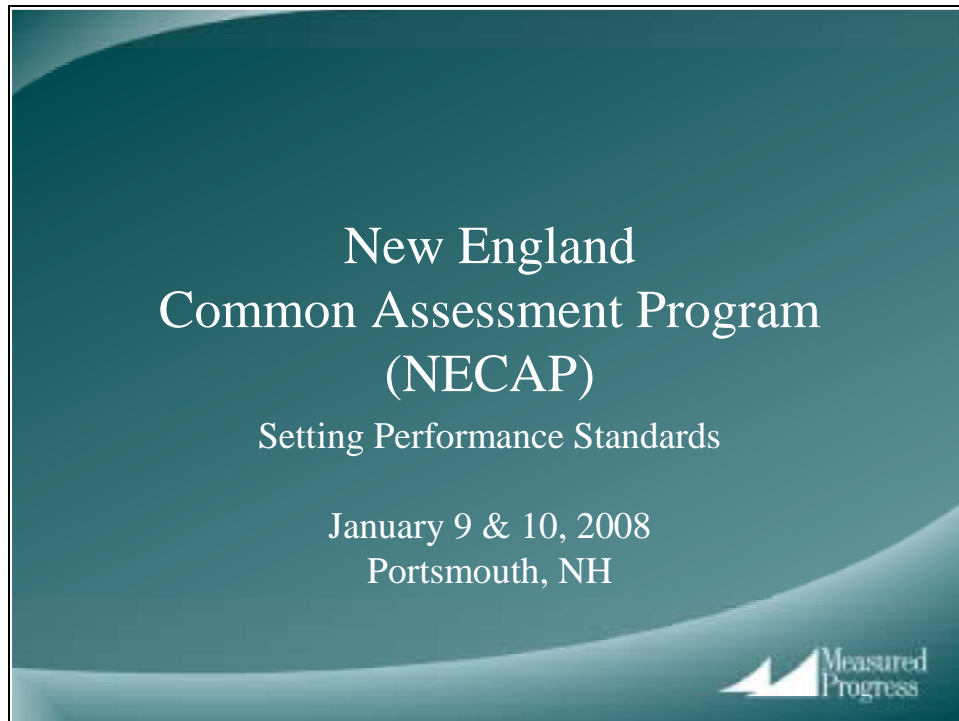
APPENDIX A: NECAP STANDARD SETTING ACHIEVEMENT LEVEL DESCRIPTIONS (ALDs)

NECAP Grade 11 General Achievement Level Descriptions	
<i>Substantially Below Proficient</i>	<p>Students performing at this level demonstrate extensive and significant gaps in the prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the grade 9-10 GSEs.</p> <p>Additional instruction and support is necessary for these students to meet the grade 9-10 GSEs.</p>
<i>Partially Proficient</i>	<p>Students performing at this level demonstrate gaps in the knowledge and skills needed to participate and perform successfully in instructional activities aligned with the grade 9-10 GSEs.</p> <p>Additional instructional support may be necessary for these students to perform successfully in courses aligned with grade 11-12 expectations.</p>
<i>Proficient</i>	<p>Students performing at this level demonstrate minor gaps in the knowledge and skills needed to participate and perform successfully in instructional activities aligned with the grade 9-10 GSEs.</p> <p>It is likely that any gaps in the prerequisite knowledge and skills demonstrated by these students can be addressed by the classroom teacher during the course of classroom instruction aligned with grade 11-12 expectations.</p>
<i>Proficient with Distinction</i>	<p>Students performing at this level demonstrate the prerequisite knowledge and skills needed to participate and excel in instructional activities aligned with the grade 9-10 GSEs. Errors made by these students are few and minor and do not reflect gaps in prerequisite knowledge and skills.</p> <p>These students are prepared to perform successfully in classroom instruction aligned with grade 11-12 expectations.</p>

Mathematics Achievement Level Descriptions	
<i>Substantially Below Proficient</i>	Student's problem solving is often incomplete, lacks logical reasoning and accuracy, and shows little conceptual understanding in most aspects of the grade span expectations. Student is able to start some problems but computational errors and lack of conceptual understanding interfere with solving problems successfully.
<i>Partially Proficient</i>	Student's problem solving demonstrates logical reasoning and conceptual understanding in some, but not all, aspects of the grade span expectations. Many problems are started correctly, but computational errors may get in the way of completing some aspects of the problem. Student uses some effective strategies. Student's work demonstrates that he or she is generally stronger with concrete than abstract situations.
<i>Proficient</i>	Student's problem solving demonstrates logical reasoning with appropriate explanations that include both words and proper mathematical notation. Student uses a variety of strategies that are often systematic. Computational errors do not interfere with communicating understanding. Student demonstrates conceptual understanding of most aspects of the grade span expectations.
<i>Proficient with Distinction</i>	Student's problem solving demonstrates logical reasoning with strong explanations that include both words and proper mathematical notation. Student's work exhibits a high level of accuracy, effective use of a variety of strategies, and an understanding of mathematical concepts within and across grade span expectations. Student demonstrates the ability to move from concrete to abstract representations.

Reading Achievement Level Descriptions	
<i>Substantially Below Proficient</i>	Student's performance demonstrates minimal ability to derive/construct meaning from grade-appropriate text. Student may be able to recognize story elements and text features. Student's limited vocabulary knowledge and use of strategies impacts the ability to read and comprehend text.
<i>Partially Proficient</i>	Student's performance demonstrates an inconsistent ability to read and comprehend grade-appropriate text. Student attempts to analyze and interpret literary and informational text. Student may make and/or support assertions by referencing text. Student's vocabulary knowledge and use of strategies may be limited and may impact the ability to read and comprehend text.
<i>Proficient</i>	Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student makes and supports relevant assertions by referencing text. Student uses vocabulary strategies and breadth of vocabulary knowledge to read and comprehend text.
<i>Proficient with Distinction</i>	Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student offers insightful observations/assertions that are well supported by references to the text. Student uses range of vocabulary strategies and breadth of vocabulary knowledge to read and comprehend a wide variety of texts.

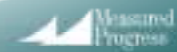
APPENDIX B: NECAP STANDARD SETTING OPENING SESSION POWERPOINT



What is Standard Setting?

- Set of activities that result in the determination of threshold or cut scores on an assessment
- We are trying to answer the question:
 - How much is enough?

3



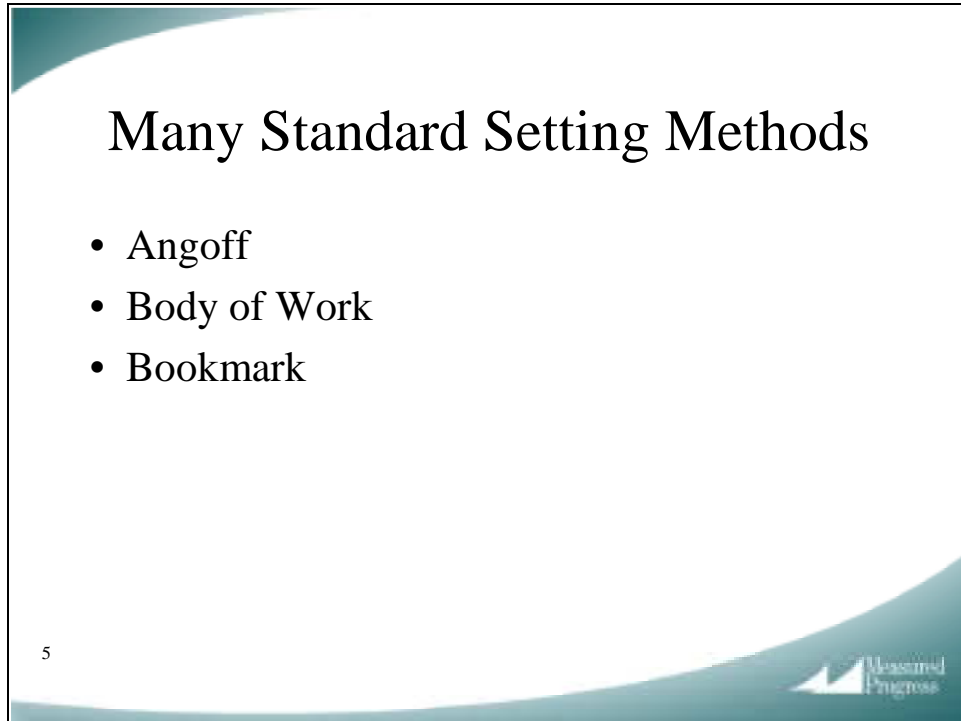
Slide 4

What is Standard Setting

- Data collection phase
 - Collection of other performance data
 - Standard-setting meeting
- Policy/Decision making phase
 - Review of data collected and final decision about placement of cut points

4



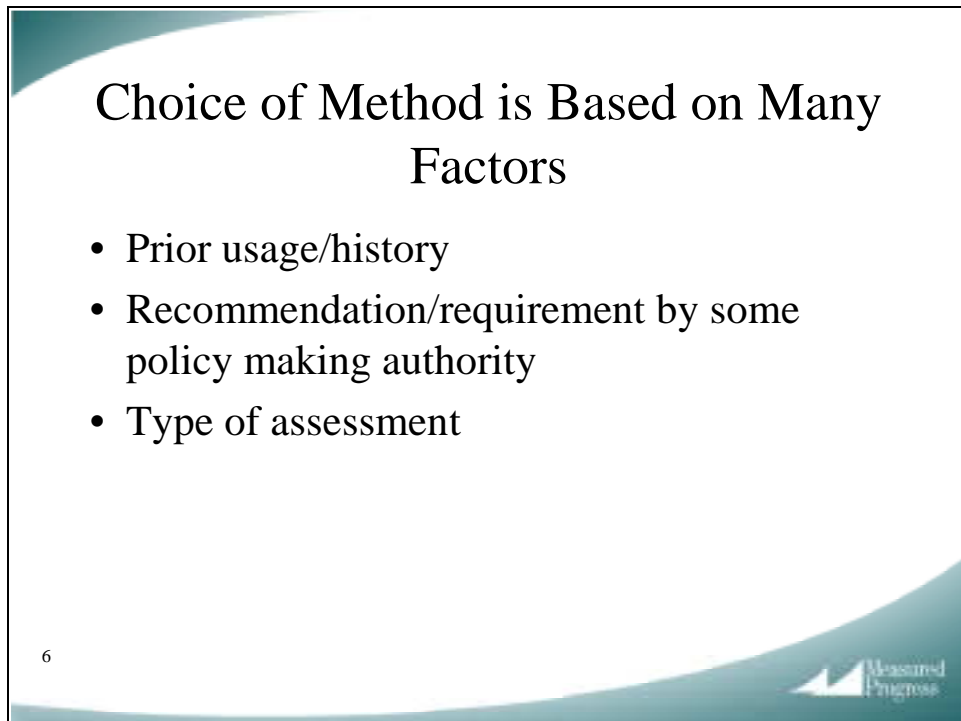
A presentation slide with a white background and a teal decorative border at the top and bottom. The title "Many Standard Setting Methods" is centered at the top in a black serif font. Below the title is a bulleted list of three items: "Angoff", "Body of Work", and "Bookmark". In the bottom left corner, the number "5" is displayed. In the bottom right corner, there is a logo consisting of a white sailboat icon and the text "Measured Progress" in a small, light blue font.

Many Standard Setting Methods

- Angoff
- Body of Work
- Bookmark

5

Measured Progress

A presentation slide with a white background and a teal decorative border at the top and bottom. The title "Choice of Method is Based on Many Factors" is centered at the top in a black serif font. Below the title is a bulleted list of three items: "Prior usage/history", "Recommendation/requirement by some policy making authority", and "Type of assessment". In the bottom left corner, the number "6" is displayed. In the bottom right corner, there is a logo consisting of a white sailboat icon and the text "Measured Progress" in a small, light blue font.

Choice of Method is Based on Many Factors

- Prior usage/history
- Recommendation/requirement by some policy making authority
- Type of assessment


6

Measured Progress

Choice of Method is Based on Many Factors

- Weighing all these factors, it was determined that the methods to be used are:
 - Reading & Math: Bookmark Method
 - Writing: modified Body of Work Method
- Both Bookmark and Body of Work are well-established procedures that have been successfully used on many assessments
- Both have produced defensible results


7



Choice of Method is Based on Many Factors

- Bookmark is appropriate for assessments that consist primarily of multiple-choice items but also include some constructed-response items
- Body of Work method works well for assessments that consist primarily or entirely of constructed-response items

8



A presentation slide with a white background and a teal decorative border at the top and bottom. The title "What Next?" is centered at the top. Below it is a bulleted list. The number "9" is in the bottom left corner, and the "Measured Progress" logo is in the bottom right corner.

What Next?

- Writing group will move to their breakout room; Reading and Math groups will stay here (for now)
 - Writing group will receive specific training on BOW method
 - Reading & Math groups will receive training on Bookmark method

9

Measured Progress

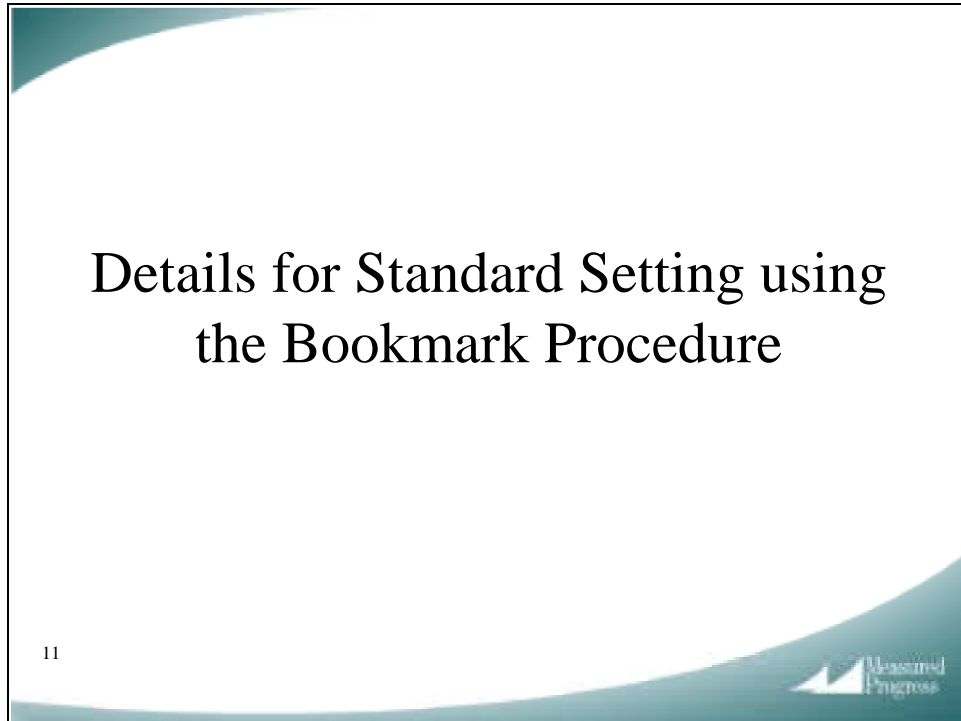
A presentation slide with a white background and a teal decorative border at the top and bottom. The title "What Next?" is centered at the top. Below it is a bulleted list. The number "10" is in the bottom left corner, and the "Measured Progress" logo is in the bottom right corner.

What Next?

- Then, Reading and Math groups will move to their breakout rooms, and all three groups will begin standard-setting activities:
 - Take the test
 - Complete the item map (Reading & Math)
 - Discuss Achievement Level Descriptions (Reading & Math) or Rubric (Writing)
 - Do ratings
 - Complete evaluation

10

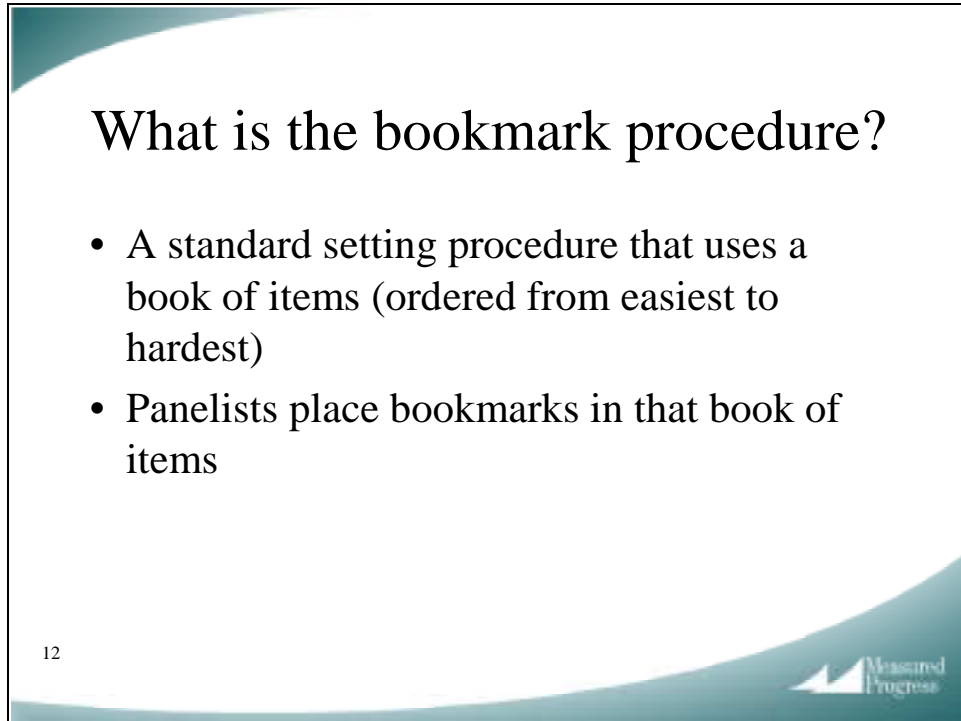
Measured Progress

The slide features a white background with a teal gradient at the top-left and bottom-right corners. The title is centered in a large, black, serif font. The slide number '11' is in the bottom-left, and the 'Measured Progress' logo is in the bottom-right.

Details for Standard Setting using the Bookmark Procedure

11

Measured Progress

The slide features a white background with a teal gradient at the top-left and bottom-right corners. The title is centered in a large, black, serif font. Below the title is a bulleted list. The slide number '12' is in the bottom-left, and the 'Measured Progress' logo is in the bottom-right.

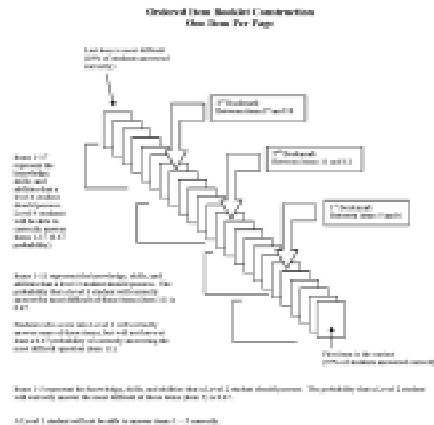
What is the bookmark procedure?

- A standard setting procedure that uses a book of items (ordered from easiest to hardest)
- Panelists place bookmarks in that book of items

12

Measured Progress

What is the bookmark procedure?



13

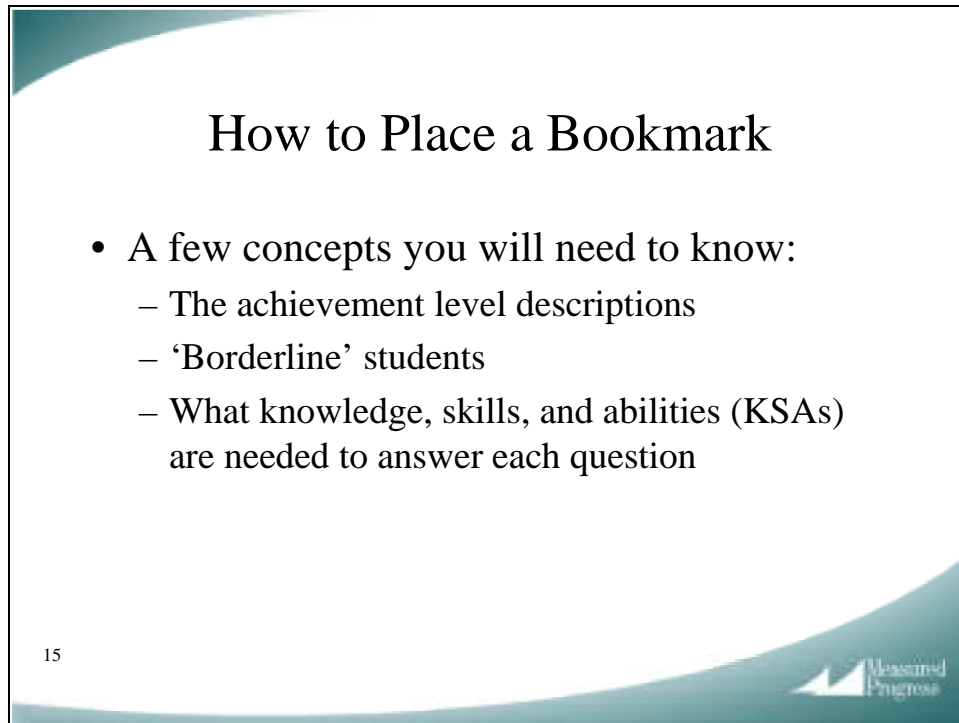
BASED ON 4. HYPOTHETICAL SITUATIONS
EXAMPLE ONLY – THESE ARE NOT INTERPRETATIONS AS ACTUAL
CITY SCORES

Unsettled Progress

A Technical Detail regarding the Bookmark Procedure

- What you need to know is that **the ordered item cut point for a given cut does not equal the raw score a student must obtain to be categorized into the higher achievement level**
- For example, if the *Substantially Below Proficient/Partially Proficient* cut is set between ordered items 3 and 4, this **does not** mean that a student only needs to get 4 points on the test in order to be classified into the *Partially Proficient* level

14

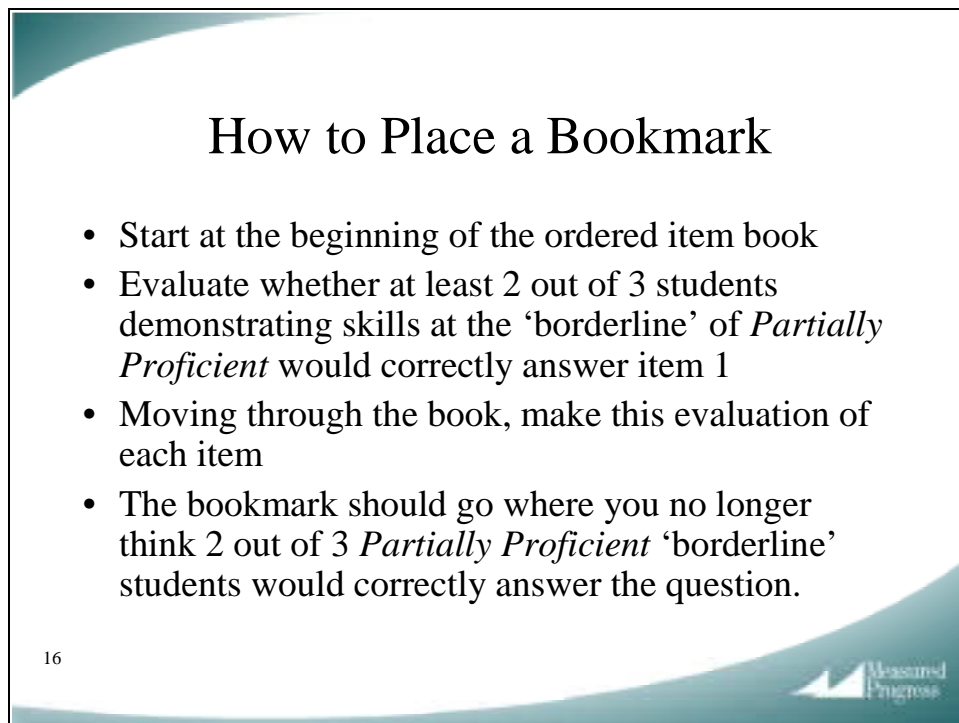


How to Place a Bookmark

- A few concepts you will need to know:
 - The achievement level descriptions
 - ‘Borderline’ students
 - What knowledge, skills, and abilities (KSAs) are needed to answer each question

15

Measured Progress



How to Place a Bookmark

- Start at the beginning of the ordered item book
- Evaluate whether at least 2 out of 3 students demonstrating skills at the ‘borderline’ of *Partially Proficient* would correctly answer item 1
- Moving through the book, make this evaluation of each item
- The bookmark should go where you no longer think 2 out of 3 *Partially Proficient* ‘borderline’ students would correctly answer the question.


16

Measured Progress

How to Place a Bookmark

Item Number	Would at least 2 out of 3 students who demonstrate skills at the <i>Partially Proficient</i> 'borderline' correctly answer this question?
1	Yes
2	Yes
3	Yes
4	Yes
5	Yes
6	Yes
7	Yes
8	Yes
9	No
10	No
11	No
12	No
13	No
14	No
15	No
...	No


17



How to Place a Bookmark

- In the example, the bookmark would go between items 8 and 9
- However, it won't be that easy; there will be gray areas
- You will have the opportunity to discuss your bookmark placements and change them if desired
- Place one bookmark for each cut score

18



How to Place a Bookmark

- To place your bookmarks you will need to be familiar with the achievement level descriptions and the assessment items

19

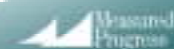


Slide 20

How to Place a Bookmark

- Don't worry, we have procedures, materials and staff to assist you in this process.

20



Slide 21

Any questions about the Bookmark Procedure?

21

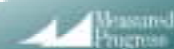


Slide 22

What Next?

- After this session, you will break into grade-level groups, where you will:
 - take the assessment to familiarize yourself with the test items
 - discuss the Achievement Level Descriptions and develop definitions of “borderline” *Partially Proficient*, *Proficient*, and *Proficient with Distinction* students

22



Slide 23

What Next?

- You will:
 - complete the Item Map, which is a document that will help you with the bookmark placement process, and
 - do the rounds of ratings

23



Slide 24

What Next?

- As the final step, we will ask you to complete an evaluation of the standard setting process

24



Slide 25

Good Luck!

25



APPENDIX C: NECAP STANDARD SETTING GENERAL INSTRUCTIONS FOR GROUP FACILITATORS - READING -

GENERAL INSTRUCTIONS FOR NECAP STANDARD SETTING GROUP FACILITATORS

READING GRADE 11

- Prior to Round 1 Ratings

Introductions:

1. Welcome group, introduce yourself (name, affiliation, a little selected background information).
2. Have each participant introduce him/herself.

- Take the Test

Overview: In order to establish an understanding of the NECAP test items and for panelists to gain an understanding of the experience of the students who take the test, each participant will take the test. Panelists may wish to discuss or take issue with the items in the test. Tell them we will gladly take their feedback to the DOE. However, this is the actual assessment that students took and it is the set of items on which we must set standards.

Activities:

- 1) Introduce NECAP and convey/do each of the following:
 - a. Tell panelists that they are about to take the actual NECAP assessment;
 - b. The purpose of the exercise is to help them establish a good understanding of the test items and to gain an understanding of the experience of the students who take the assessment;
- 2) Give each panelist a test booklet;
- 3) Tell panelists to try to take on the perspective of a student as they complete the test.
- 4) When the majority of the panelists have finished, pass out answer key

Fill Out Item Map

Overview: The primary purpose of filling out the item map is for panelists to think about and document the knowledge, skills, and abilities students need to answer each question. Panelists should have an understanding of what makes one test item harder or easier than another. The notes panelists take here will be useful in helping them place their bookmarks and in discussions during the two rounds of ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. Item map
 - b. Ordered item book
2. Review the ordered item book and item map with the panelists. Explain what each is, and point out the correspondence of the ordered items between the two. Explain that the items are ordered from easiest to hardest, and that items worth more than 1 point will appear once for each possible score point.
3. Provide an overview of the task paraphrasing the following:
 - a. The primary purpose of this activity is for panelists to think about what makes one question harder or easier than another. For example, it may be that the concept tested is a difficult concept, or that the concept isn't difficult but that the particular wording of the question makes it a difficult question. Similarly, the concept may be a difficult one, but the wording of the question makes it easier.
 - b. Panelists should take notes about their thoughts regarding each question. These will be useful in the rating activities and later discussions.
4. Tell panelists to work individually at first. After they complete the item map they will have the opportunity to discuss it with their colleagues.
5. Note that, for the bottom cut, panelists will begin the item mapping process with the first ordered item. For the remaining two cuts, they should start five ordered items before the starting cut.
6. Each panelist will begin with the first ordered item and compare it to the next ordered item. What makes the second item harder than the first? Panelists should not agonize over these decisions. It may be that the second item is only slightly harder than the first.
7. Panelists should work their way through the item map, stopping about five ordered items after the *Substantially Below Proficient/Partially Proficient* starting cut.
8. Panelists will then do the same process for the *Partially Proficient/Proficient* and *Proficient/Proficient with Distinction* cuts, each time starting approximately five ordered items before the cut and ending approximately five ordered items after the cut.
9. Note that panelists may feel that they need to expand the range of items they consider in one direction or the other. Five ordered items before and after the starting cuts is a guideline, but they may consider more items if necessary.
10. Once panelists have completed the item map, they should discuss them as a group.
11. Based on the group discussion, the panelists should modify their own item map (make additional notes, cross things out, etc...)

Discuss Achievement Level Descriptions & Describe Characteristics of the “Borderline” Student

Overview: In order to establish an understanding of the expected performance of borderline students

on the test, panelists must have a clear understanding of:

- 1) The definition of the four achievement levels, and
- 2) Characteristics of students who are “just able enough” to be classified into each achievement level. These students will be referred to as borderline students, since they are right on the border between achievement levels.

The purpose of this activity is for the panelists to obtain an understanding of the Achievement Level Descriptions with an emphasis on characteristics that describe students at the borderline -- both what these students can and cannot do.

This activity is critical since the ratings panelists will be making in Rounds 1 and 2 will be based on these understandings.

Activities:

- 1) Introduce task. In this activity they will:
 - a. Individually review the Achievement Level Descriptions;
 - b. discuss Descriptions as a group; and
 - c. generate whole group descriptions of borderline *Partially Proficient*, *Proficient* and *Proficient with Distinction* students.The facilitator should compile the descriptions as bulleted lists on chart paper; the chart paper will then be posted so the panelists can refer to the lists as they go through the bookmark process.
- 2) Pass out the Achievement Level Descriptions and have panelists individually review them. Panelists can make notes if they like.
- 3) After individually reviewing the Descriptions, have panelists discuss each one as a whole group, starting with *Partially Proficient*, and provide clarification. The goal here is for the panelists to have a collegial discussion in which to bring up/clarify any issues or questions, and to come to a common understanding of what it means to be in each achievement level. It is not unusual for panelists to disagree with the descriptions they will see; almost certainly there will be some panelists who will want to change them. However, the task at hand is for panelists to have a common understanding of what knowledge, skills, and abilities are described by each Achievement Level Description.
- 4) Once panelists have a solid understanding of the Achievement Level Descriptions, have them focus their discussion on the knowledge, skills, and abilities of students who are in the *Partially Proficient* category, but just barely. The focus should be on those characteristics and KSAs that best describe the lowest level of performance necessary to warrant a *Partially Proficient* classification.
- 5) After discussing *Partially Proficient*, have the panelists discuss characteristics of the borderline *Proficient* student and then characteristics of the borderline *Proficient with Distinction* student. Panelists should be made aware of the importance of the *Proficient* cut.
- 6) Using chart paper, generate a bulleted list of characteristics for each of the levels based on the entire room discussion. Post these on the wall of the room.

▪ Round 1

Overview of Round 1: The primary purpose of Round 1 is to ask the panelists to evaluate and, if necessary, revise the starting cut points. For this round, panelists will work as a group. Beginning

with the starting *Substantially Below Proficient/Partially Proficient* cut point, panelists will evaluate each item, starting approximately five ordered items before the cut and ending approximately five ordered items after the cut (or as appropriate). (Note, again, that panelists may feel that they need to expand the range of items they consider in one direction or the other. Five ordered items before and after the starting cuts is a guideline, but they may consider more items if necessary.) The panelists will gauge the level of difficulty of each of the items for those students who barely meet the definition of *Partially Proficient*. The task that panelists are asked to do is to estimate whether a borderline *Partially Proficient* student would answer each question correctly. More specifically panelists should answer:

- Would *at least* 2 out of 3 students performing at the borderline answer the question correctly?

In the case of open-response questions, panelists should ask:

- Would *at least* 2 out of 3 students performing at the borderline get this score point *or higher*?

This same process is then repeated for the starting *Partially Proficient/Proficient* cut and the starting *Proficient/Proficient with Distinction* cut.

Activities:

1. Make sure panelists have the following materials:
 - a. Round 1 rating form
 - b. Ordered Item Book
 - c. Item Map
 - d. Achievement Level Descriptions
 - e. Starting cut points
2. Have panelists write round number 1 and their ID number on the rating form. The ID number is on their name tags.
3. Provide an overview of Round 1. Paraphrase the following:
 - a. Orient panelists to the ordered-item book. Explain that the items are ordered from easiest to hardest; for constructed-response items, explain that each item appears once for each possible score point.
 - b. Orient panelists to the starting cut points. Make sure panelists understand that the ordered item cut point for *SBP/PP* is not the same as the raw score a student must obtain in order to be classified into *Partially Proficient*. For example, if a starting cut point is between ordered items 6 and 7, that **does not** mean that a student only needs 7 points to be classified as *Partially Proficient*.
 - c. The primary purpose of this activity is for the panelists to discuss whether students whose performance is barely *Partially Proficient* would correctly answer each item, beginning approximately five positions prior to the starting *Substantially Below Proficient/Partially Proficient* cut, and to place their bookmark where they believe the answer of ‘yes’ turns to ‘no’. Remind panelists that they should be thinking about two-thirds of the borderline students. Once they have completed the process for the *Substantially Below Proficient/Partially Proficient* cut, they will proceed to the remaining two cut points.
 - d. Each panelist needs to base their judgments on his/her experience with the content, understanding of students, and the definitions of the borderline students generated previously.
 - e. One bookmark will be placed for each cut point.

- f. If panelists are struggling with placing a particular bookmark they should use their best judgment and move on. They will have an opportunity to revise their ratings.
 - g. Panelists should feel free to take notes if there are particular points about where they placed their bookmarks that they think are worthy of discussion in Round 2.
4. Tell panelists that they will be discussing each cut point with the other panelists, but that they will be placing the bookmarks individually. It is not necessary for the panelists to come to consensus about whether and how the cut points should be revised.
5. Go over the rating form with panelists.
 - a. Lead panelists through a step-by-step demonstration of how to fill in the rating form.
 - b. Answer questions the panelists may have about the work in Round 1.
 - c. Once everyone understands what they are to do in Round 1, tell them to begin.
6. Using the ordered item book, the panelists begin approximately five ordered items prior to the starting *Substantially Below Proficient/Partially Proficient* cut, or as appropriate.
7. After they have placed the first bookmark, they will proceed to the *Partially Proficient/Proficient* cut, beginning approximately five ordered items prior to the starting cut.
8. After they have placed the second bookmark, they will proceed to the *Proficient/Proficient with Distinction* cut, again beginning approximately five ordered items prior to the starting cut.
9. After they have placed all three bookmarks, have panelists fill out their rating forms. Ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. **The round number and ID number must be filled in.**
 - b. **The item numbers identifying each cut score must be adjacent.**
 - c. Check each panelist's rating form before you allow them to leave for a short break.
 - d. When all the rating forms have been collected, the group will take a break. Immediately bring the rating forms to the R&A work room for tabulation.

▪ Tabulation of Round 1 Results

Tabulation of Round 1 results will be completed as quickly as possible after receipt of the rating forms.

▪ Round 2

Overview of Round 2: The primary purpose of Round 2 is to ask the panelists to discuss their Round 1 placements as a whole group and to revise their ratings on the basis of that discussion. They will discuss their ratings in the context of the ratings made by other members of the group. The panelists with the highest and lowest ratings should comment on why they gave the ratings they did. The group should get a sense of how much variation there is in the ratings. Panelists should also consider the question, “How tough or easy a panelist are you?” The purpose here is to allow panelists to examine their individual expectations (in terms of their experiences) and to share these expectations and experiences in order to attain a better understanding of how their experiences impact their decision-making.

To aid with the discussion, panelists will also be given impact data, showing the approximate percentage of students who would be classified into each achievement level category based on the room average bookmark placements from Round 1.

Once panelists have reviewed and discussed their bookmark placements, they will be given the opportunity to change or revise their Round 1 ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. The Round 2 rating forms
 - b. Ordered item booklets
 - c. Item maps
 - d. Achievement Level Descriptions
2. Have panelists write round number 2 and their ID number on the rating form.
3. A psychometrician will present and explain the following information to the panelists:
 - a. The average bookmark placement for the whole group based on the Round 1 ratings. Based on their Round 1 ratings, panelists will know where they fall relative to the group average.
 - b. Impact data, showing the approximate percentage of students across the three states that would be classified into each achievement level category based on the room average bookmark placements from Round 1.
4. Provide an overview of Round 2. Paraphrase the following:
 - a. As in Round 1, the primary purpose is to place bookmarks where you feel the achievement levels are best distinguished, considering the additional information and further discussion.
 - b. Each panelist needs to base his/her judgments on his/her experience with the content area, understanding of students, the definitions of the borderline students generated previously, discussions with other panelists and the knowledge, skills, and abilities required to answer each item.
5. Panelists should be given a few minutes to review the Round 1 average cut points and impact data.
6. Once they have reviewed the information, the panelists will discuss their Round 1 ratings, beginning with the first cut point.
 - a. The discussion should focus on differences in where individual panelists placed their cutpoints.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.

- d. On the basis of the discussions and the feedback presented, panelists should make a second round of ratings.
- e. When placing their Round 2 bookmarks, panelists should not feel compelled to change their ratings.
- f. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is consistently higher or lower than the group, they may have a different understanding of the borderline student than the rest of the group, or a different understanding of the Achievement Level Descriptions, or both. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Achievement Level Descriptions.**

- 7. When the group has completed their second ratings, collect the rating forms. When you collect the rating forms carefully inspect them to ensure they are filled out properly.
 - a. **The round number and panelist ID number must be filled in.**
 - b. **The item numbers identifying each cut score must be adjacent.**
 - c. Check each panelist's rating form before you allow them to leave for a short break.
 - d. When all the rating forms have been collected, the group will take a break. Immediately bring the rating forms to the R&A work room for tabulation.

- Complete Evaluation Form

Upon completion of Round 2, have panelists fill out the evaluation form. Emphasize that their honest feedback is important.

APPENDIX D: NECAP STANDARD SETTING GENERAL DIRECTIONS FOR GROUP FACILITATORS - MATHEMATICS –

GENERAL INSTRUCTIONS FOR NECAP STANDARD SETTING GROUP FACILITATORS

MATHEMATICS GRADE 11

- Prior to Round 1 Ratings

Introductions:

3. Welcome group, introduce yourself (name, affiliation, a little selected background information).
4. Have each participant introduce him/herself.

- Take the Test

Overview: In order to establish an understanding of the NECAP test items and for panelists to gain an understanding of the experience of the students who take the test, each participant will take the test. Panelists may wish to discuss or take issue with the items in the test. Tell them we will gladly take their feedback to the DOE. However, this is the actual assessment that students took and it is the set of items on which we must set standards.

Activities:

- 5) Introduce NECAP and convey/do each of the following:
 - a. Tell panelists that they are about to take the actual NECAP assessment;
 - b. The purpose of the exercise is to help them establish a good understanding of the test items and to gain an understanding of the experience of the students who take the assessment;
- 6) Give each panelist a test booklet;
- 7) Tell panelists to try to take on the perspective of a student as they complete the test.
- 8) When the majority of the panelists have finished, pass out answer key

Fill Out Item Map

Overview: The primary purpose of filling out the item map is for panelists to think about and document the knowledge, skills, and abilities students need to answer each question. Panelists should have an understanding of what makes one test item harder or easier than another. The notes panelists take here will be useful in helping them place their bookmarks and in discussions during the three rounds of ratings.

Activities:

12. Make sure panelists have the following materials:
 - a. Item map
 - b. Ordered item book
13. Review the ordered item book and item map with the panelists. Explain what each is, and point out the correspondence of the ordered items between the two. Explain that the items are ordered from easiest to hardest, and that items worth more than 1 point will appear once for each possible score point.
14. Provide an overview of the task paraphrasing the following:
 - a. The primary purpose of this activity is for panelists to think about what makes one question harder or easier than another. For example, it may be that the concept tested is a difficult concept, or that the concept isn't difficult but that the

- particular wording of the question makes it a difficult question. Similarly, the concept may be a difficult one, but the wording of the question makes it easier.
- b. Panelists should take notes about their thoughts regarding each question. These will be useful in the rating activities and later discussions.
15. Tell panelists to work individually at first. After they complete the item map they will have the opportunity to discuss with their colleagues.
 16. Each panelist will begin with the ordered item number one and compare it to the next ordered item. What makes the second item harder than the first? Panelists should not agonize over these decisions. It may be that the second item is only slightly harder than the first.
 17. Panelists will continue this process, working their way through the item map and ordered item booklet.
 18. Once panelists have completed their item maps, they should discuss them as a group.
 19. Based on the group discussion, the panelists should modify their own item map (make additional notes, cross things out, etc...)

Discuss Achievement Level Descriptions & Describe Characteristics of the “Borderline” Student

Overview: In order to establish an understanding of the expected performance of borderline students on the test, panelists must have a clear understanding of:

- 3) The definition of the four achievement levels, and
- 4) Characteristics of students who are “just able enough” to be classified into each achievement level. These students will be referred to as borderline students, since they are right on the border between achievement levels.

The purpose of this activity is for the panelists to obtain an understanding of the Achievement Level Descriptions with an emphasis on characteristics that describe students at the borderline -- both what these students can and cannot do.

This activity is critical since the ratings panelists will be making in Rounds 1 through 3 will be based on these understandings.

Activities:

- 7) Introduce task. In this activity they will:
 - a. Individually review the Achievement Level Descriptions;
 - b. discuss Descriptions as a group; and
 - c. generate whole group descriptions of borderline *Partially Proficient*, *Proficient* and *Proficient with Distinction* students.

The facilitator should compile the descriptions as bulleted lists on chart paper; the chart paper will then be posted so the panelists can refer to the lists as they go through the bookmark process.
- 8) Pass out the Achievement Level Descriptions and have panelists individually review them. Panelists can make notes if they like.
- 9) After individually reviewing the Descriptions, have panelists discuss each one as a whole group, starting with *Partially Proficient*, and provide clarification. The goal here is for

- the panelists to have a collegial discussion in which to bring up/clarify any issues or questions, and to come to a common understanding of what it means to be in each achievement level. It is not unusual for panelists to disagree with the descriptions they will see; almost certainly there will be some panelists who will want to change them. However, the task at hand is for panelists to have a common understanding of what knowledge, skills, and abilities are described by each Achievement Level Description.
- 10) Once panelists have a solid understanding of the Achievement Level Descriptions, have them focus their discussion on the knowledge, skills, and abilities of students who are in the *Partially Proficient* category, but just barely. The focus should be on those characteristics and KSAs that best describe the lowest level of performance necessary to warrant a *Partially Proficient* classification.
 - 11) After discussing *Partially Proficient*, have the panelists discuss characteristics of the borderline *Proficient* student and then characteristics of the borderline *Proficient with Distinction* student. Panelists should be made aware of the importance of the *Proficient* cut.
 - 12) Using chart paper, generate a bulleted list of characteristics for each of the levels based on the entire room discussion. Post these on the wall of the room.

▪ Round 1

Overview of Round 1: The purpose of Round 1 is for the panelists to make their initial judgments as to where the bookmarks should be placed. For this round, panelists will work individually, without consulting with their colleagues. Beginning with the first ordered item and the *Substantially Below Proficient/Partially Proficient* cut point, panelists will evaluate each item in turn. The panelists will gauge the level of difficulty of each of the items for those students who barely meet the definition of *Partially Proficient*. The task that panelists are asked to do is to estimate whether a borderline *Partially Proficient* student would answer each question correctly. More specifically panelists should answer:

- Would *at least* 2 out of 3 students performing at the borderline answer the question correctly?

In the case of open-response questions, panelists should ask:

- Would *at least* 2 out of 3 students performing at the borderline get this score point *or higher*?

This same process is then repeated for the starting *Partially Proficient/Proficient* cut and the starting *Proficient/Proficient with Distinction* cut.

Activities:

10. Make sure panelists have the following materials:
 - a. Round 1 rating form
 - b. Ordered Item Book
 - c. Item Map
 - d. Achievement Level Descriptions
11. Have panelists write round number 1 and their ID number on the rating form. The ID number is on their name tags.
12. Provide an overview of Round 1, covering each of the following:

- a. Orient panelists to the ordered-item book. Explain that the items are ordered from easiest to hardest; for open-response items, explain that each item appears once for each possible score point.
 - b. The primary purpose of this activity is for the panelists to discuss whether students whose performance is barely *Partially Proficient* would correctly answer each item, and to place their bookmark where they believe the answer of ‘yes’ turns to ‘no’. Remind panelists that they should be thinking about two-thirds of the borderline students. Once they have completed the process for the *Substantially Below Proficient/Partially Proficient* cut, they will proceed to the remaining two cut points.
 - c. Each panelist needs to base his/her judgments on his/her experience with the content, understanding of students, and the definitions of the borderline students generated previously.
 - d. One bookmark will be placed for each cut point.
 - e. If panelists are struggling with placing a particular bookmark they should use their best judgment and move on. They will have an opportunity to revise their ratings in Rounds 2 and 3.
 - f. Panelists should feel free to take notes if there are particular points about where they placed their bookmarks that they think are worthy of discussion in Round 2.
13. Go over the rating form with panelists.
 - a. Lead panelists through a step-by-step demonstration of how to fill in the rating form.
 - b. Answer questions the panelists may have about the work in Round 1.
 - c. Once everyone understands what they are to do in Round 1, tell them to begin.
 14. The panelists begin with ordered item number 1 and proceed through the ordered item booklet, each time asking whether at least two out of three borderline students would correctly answer the question. They will place their first bookmark at the point where the answer changes from “yes” to “no.”
 15. After they have placed the first bookmark, they will continue through the ordered item booklet, making the same judgments for the *Partially Proficient/Proficient* cut, and the *Proficient/Proficient with Distinction* cut.
 16. After they have placed all three bookmarks, have panelists fill out their rating forms. Ask them to carefully inspect their rating forms to ensure they are filled out properly.
 - a. **The round number and ID number must be filled in.**
 - b. **The item numbers identifying each cut score must be adjacent.**
 - c. Check each panelist’s rating form before you allow them to leave for a short break.
 - d. When all the rating forms have been collected, the group will take a break. Immediately bring the rating forms to the R&A work room for tabulation.

▪ **Tabulation of Round 1 Results**

Tabulation of Round 1 results will be completed as quickly as possible after receipt of the rating forms.

▪ **Round 2**

Overview of Round 2: The primary purpose of Round 2 is to ask the panelists to discuss their Round 1 placements as a whole group and to revise their ratings on the basis of that discussion. They will discuss their ratings in the context of the ratings made by other members of the group. The panelists with the highest and lowest ratings should comment on why they gave the ratings they did. The group should get a sense of how much variation there is in the ratings. Panelists should also consider the question, “How tough or easy a panelist are you?” The purpose here is to allow panelists to examine their individual expectations (in terms of their experiences) and to share these expectations and experiences in order to attain a better understanding of how their experiences

impact their decision-making.

To aid with the discussion, a psychometrician will present the group with the room average bookmark placements from Round 1.

Once panelists have reviewed and discussed their bookmark placements, they will be given the opportunity to change or revise their Round 1 ratings.

Activities:

8. Make sure panelists have the following materials:
 - a. The Round 2 rating forms
 - b. Ordered item booklets
 - c. Item maps
 - d. Achievement Level Descriptions
9. Have panelists write round number 2 and their ID number on the rating form.
10. A psychometrician will present and explain the average bookmark placement for the whole group based on the Round 1 ratings. Based on their Round 1 ratings, panelists will know where they fall relative to the group average. This information is useful so that panelists get a sense if they are more stringent or more lenient than other panelists.
11. Provide an overview of Round 2. Paraphrase the following:
 - a. As in Round 1, the primary purpose is to place bookmarks where you feel the achievement levels are best distinguished, considering the additional information and further discussion.
 - b. Each panelist needs to base his/her judgments on his/her experience with the content area, understanding of students, the definitions of the borderline students generated previously, discussions with other panelists and the knowledge, skills, and abilities required to answer each item.
12. Panelists should be given a few minutes to review the bookmark placements based on the room average cut points from Round 1.
13. Once they have reviewed the information, the panelists will discuss their Round 1 ratings, beginning with the first cut point.
 - g. The discussion should focus on differences in where individual panelists placed their cutpoints.
 - h. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - i. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - j. On the basis of the discussions and the feedback presented, panelists should make a second round of ratings.
 - k. When placing their Round 2 bookmarks, panelists should not feel compelled to change their ratings.
 - l. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is consistently higher or lower than the group, they may have a different understanding of the borderline student than the rest of the group, or a different understanding of the Achievement Level Descriptions, or both. **It is O.K. for panelists to disagree, but that disagreement should be based on a common**

understanding of the Achievement Level Descriptions.

14. When the group has completed their second ratings, collect the rating forms. When you collect the rating forms carefully inspect them to ensure they are filled out properly.
 - a. **The round number and panelist ID number must be filled in.**
 - b. **The item numbers identifying each cut score must be adjacent.**
 - c. Check each panelist's rating form before you allow them to leave for a short break.
 - d. When all the rating forms have been collected, the group will take a break.
Immediately bring the rating forms to the R&A work room for tabulation.

- Tabulation of Round 2 Results

Round 2 results will be tabulated as soon as possible upon receipt of the rating forms.

- Round 3

Overview of Round 3: In Round 3, the panelists will discuss their Round 2 ratings as a whole group and have another opportunity to revise their ratings on the basis of that discussion. Again, they will discuss their ratings in the context of the ratings made by other members of the group.

To aid with the discussion, a psychometrician **may** present impact data to each group (this decision will be made at the standard-setting meeting). The impact data shows the approximate percentage of students across the three states that would be classified into each achievement level category based on the room average bookmark placements from Round 2.

Once panelists have reviewed and discussed their bookmark placements, they will be given the opportunity to revise their Round 2 ratings.

Activities:

1. Make sure panelists have the following materials:
 - a. The Round 3 rating forms
 - b. Ordered item booklets
 - c. Item maps
 - d. Achievement Level Descriptions
2. Have panelists write round number 3 and their ID number on the rating form.
3. A psychometrician will present and explain the average bookmark placement for the whole group based on the Round 2 ratings. Again, based on their Round 2 ratings, panelists will know where they fall relative to the group average. The psychometrician **may** also present impact data, showing the approximate percentage of students across the three states that would be classified into each achievement level category based on the room average bookmark placements from Round 2.
4. Provide an overview of Round 3. Paraphrase the following:
 - a. As in Rounds 1 and 2, the primary purpose is to place bookmarks where you feel the achievement levels are best distinguished, considering the additional information and further discussion.
 - b. Each panelist needs to base his/her judgments on his/her experience with the content area, understanding of students, the definitions of the borderline students generated

previously, discussions with other panelists and the knowledge, skills, and abilities required to answer each item.

5. Panelists should be given a few minutes to review the Round 2 average cut points and impact data (if presented).
6. Once they have reviewed the materials, the panelists will discuss their Round 2 ratings, beginning with the first cut point.
 - a. The discussion should focus on differences in where individual panelists placed their cutpoints.
 - b. Panelists should be encouraged to listen to their colleagues as well as express their own points of view.
 - c. If the panelists hear a logic/rationale/argument that they did not consider and that they feel is compelling, then they may adjust their ratings to incorporate that information.
 - d. On the basis of the discussions and the feedback presented, panelists should make a third round of ratings.
 - e. When placing their Round 3 bookmarks, panelists should not feel compelled to change their ratings.
 - f. The group does not have to achieve consensus. If panelists honestly disagree, that is fine. We are trying to get the best judgment of each panelist. Panelists should not feel compelled or coerced into making a rating they disagree with.

Encourage the panelists to use the discussion and feedback to assess how stringent or lenient a judge they are. If a panelist is consistently higher or lower than the group, they may have a different understanding of the borderline student than the rest of the group, or a different understanding of the Achievement Level Definitions, or both. **It is O.K. for panelists to disagree, but that disagreement should be based on a common understanding of the Achievement Level Definitions.**

7. When the group has completed their third round of ratings, collect the rating forms. When you collect the rating forms **carefully inspect them** to ensure they are filled out properly.
 - a. **The round number and panelist ID number must be filled in.**
 - b. **The item numbers identifying each cut score must be adjacent.**
 - c. Immediately provide the completed rating forms to R&A. The panelists will not see the results from this round.

▪ Complete Evaluation Form

Upon completion of Round 3, have panelists fill out the evaluation form. Emphasize that their honest feedback is important.

APPENDIX E: NECAP STANDARD SETTING GENERAL DIRECTIONS FOR GROUP FACILITATORS - WRITING –

NECAP Grade 11 Writing Standards Validation

Procedures for Writing Standards Validation Panel Meetings

- 1.) Introductions, purpose of standards validation panel meeting, and overview of the writing test design (Tim) **~15-30 minutes**
- 2.) Discussion of the five (5) writing scoring rubrics for grade 11. The common prompt this year is “Response to Informational Text”, the rubrics for “Response to Literary Text”, “Reflective Essay”, “Persuasive Essay”, “Report”, and “Procedure” will also be reviewed and discussed. Each rubric is essentially the same, with some bullets specific to the type of writing. (DOE content specialists) **~30 minutes**
- 3.) Discussion of common prompt anchor papers and scores assigned to the papers. Intent of this section is to ensure that panelists are comfortable with the scoring process and understand the relationship between the rubrics and student work. (DOE Content Specialists) **~45 minutes**

Note to Specialists: We’ve packaged all of the anchor papers and will let you choose the ones you want to highlight (total of 14 papers, two per score point).

- 4.) Overview and discussion of the general NECAP grade 11 Achievement Level Descriptors. These are the descriptors that were used during teacher judgment ratings. (Tim) **~15 minutes**
- 5.) Overview of achievement level definitions and cut scores. Explanation of the process and steps taken by state content specialists to arrive at the starting achievement level definitions, which will be validated by the panelists. (Tim) **~5 minutes**

Note to Specialists: This will be a very short description connecting the rubric language to the general descriptors.

- 6.) Rationale for the starting cut points and current achievement level definition based on the descriptions contained in the rubrics. The starting cuts are as follows:

- Substantially Below Proficient/Partially Proficient = 3/4,
- Partially Proficient/Proficient = 6/7,
- Proficient/Proficient with Distinction = 9/10.

Panelists will discuss the rationale for the cuts and the details of the definitions and may propose alternatives—with rationale. The common anchor papers and rubrics from Step #3 will help guide discussion. Notes of discussion will be recorded. (DOE Content Specialists) **~30-45 minutes**

Note to Specialists: This should be around lunch time.

- 7.) Panelists will take 5-7 minutes to write to the common prompt. This will help familiarize them with the prompt and understand how students may have answered. Next, panelists will examine of a set of student responses from the common prompt. This set will consist of approximately 16 papers distributed across the score range 2-12. (For example, a possible score distribution is 1 each for score points 2-4, 2 papers each for score points 5-9, and 1 paper each for score points 10-12.) The papers will be rank ordered lowest to highest, labeled with letters as an identifier, and will **not** have scores displayed.

Note to Specialists: You'll be picking these this afternoon on your call with Amanda.

- Panelists will be asked to read through the papers and place them in each of the 4 achievement levels. Does this level of work fit the description of proficiency in the definition? Panelists will use a rating form similar to one for a Body of Work standard setting.
 - The facilitator will tally the panelists' placements and present the results to the group. Group discussion will focus on disagreements regarding the achievement level for specific papers. Discussion will be focused in reference to the rubric. Group consensus is not needed.
 - The group will then be told the scores of each paper, so as not to influence their first rating and discussion. Panelists will then do a second rating of the set of papers.
 - This *should* bring the group to the end of day one. **3-3 ½ hours**
- 8.) Examination of student responses from the matrix prompts one at a time. This set of responses will consist of 11 papers across the 2-12 score range. This set will display the score for each paper. Panelists repeat step 7 for each of the 5 prompts. **~45-60 minutes per prompt**
- 9.) Final discussion at the end of the process across all prompts. Final rating of the achievement level definition (cut scores). **~30 minutes**

Selection of student work samples

Common prompt: papers for even score points will be selected from the scoring training pack since these have already been approved by the DOE. (The anchor papers will have already been used in step 3). Papers for odd score points (adjacent scores when double scored) will be identified by R&A. A selection will be made by Measured Progress and sent to the DOE content specialists for approval. The selection will include MP's recommendations and extra papers in case the DOE is in disagreement. A chart will be provided to organize the papers and facilitate discussion.

Matrix prompts: papers for even score points will be selected from the anchor pack and training pack, if necessary. Odd score points will be selected in the same way as the common prompt.

Measured Progress will send the selected odd score point papers to the DOE content specialists on January 2 (for delivery January 3). A conference call will be held on January 4th to come to agreement on which papers will be used.

Schedule

The DOE content specialists will meet at Measured Progress the afternoon of January 8th to finalize their role as facilitators. They will also meet with Tim at 7:30AM on January 9th.

January 9th: The entire group will start together at 9:00AM for a quick welcome and overview. The writing group will leave to start their work, while the reading and mathematics panelists stay for training on the bookmark method. Work will end around 4:00-4:30PM

January 10th: Work will start at 8:30AM., and the day should conclude by 4:00PM.

APPENDIX F: NECAP STANDARD SETTING GRADE 11 RATING FORM - READING/MATHEMATICS -

NECAP Reading Grade 11 Rating Form

Round _____

ID _____

Substantially Below Proficient Ordered Item Numbers	Partially Proficient Ordered Item Numbers	Proficient Ordered Item Numbers	Proficient with Distinction Ordered Item Numbers
First 1	First	First	First
Last _____	Last _____	Last _____	Last 52

Directions: Please enter the range of ordered item numbers that fall into each achievement level category according to where you placed your cutpoints.

Note: The ranges must be adjacent to each other. For example: Substantially Below Proficient: 1-13, Partially Proficient: 14-26, Proficient: 27-39, Proficient with Distinction: 40-52.

NECAP Mathematics Grade 11 Rating Form

Round _____

ID _____

Substantially Below Proficient Ordered Item Numbers	Partially Proficient Ordered Item Numbers	Proficient Ordered Item Numbers	Proficient with Distinction Ordered Item Numbers
First 1 Last _____	First _____ Last _____	First _____ Last _____	First _____ Last 64

Directions: Please enter the range of ordered item numbers that fall into each achievement level category according to where you placed your cutpoints.

Note: The ranges must be adjacent to each other. For example: Substantially Below Proficient: 1-16, Partially Proficient: 17-32, Proficient: 33-48, Proficient with Distinction: 49-64.

APPENDIX G: NECAP GRADE 11 FINAL WRITING RUBRICS

Grade 11 Writing Rubric - Response to Literary or Informational Text

6
<ul style="list-style-type: none"> • purpose is clear throughout; strong focus/controlling idea OR strongly stated purpose focuses the writing • intentionally organized for effect • fully developed details; rich and/or insightful elaboration supports purpose • distinctive voice, tone, and style enhance meaning • consistent application of the rules of grade-level grammar, usage, and mechanics
5
<ul style="list-style-type: none"> • purpose is clear; focus/controlling idea is maintained throughout • well-organized and coherent throughout • details are relevant and support purpose; details are sufficiently/purposely elaborated • strong command of sentence structure; uses language to enhance meaning • consistent application of the rules of grade-level grammar, usage, and mechanics
4
<ul style="list-style-type: none"> • purpose is evident; focus/controlling idea may not be maintained • generally organized and coherent • details are generally relevant and appropriately developed • well-constructed sentences; uses language well • may have some errors in grammar, usage, and mechanics
3
<ul style="list-style-type: none"> • writing has a general purpose • some sense of organization; may have lapses in coherence • some relevant details support purpose • uses language adequately; may show little variety of sentence structures • may have some errors in grammar, usage, and mechanics
2
<ul style="list-style-type: none"> • attempted or vague purpose • attempted organization; lapses in coherence • generalized, listed, or undeveloped details • may lack sentence control or may use language poorly • may have errors in grammar, usage, and mechanics that interfere with meaning
1
<ul style="list-style-type: none"> • minimal evidence of purpose • little or no organization • minimal or random information • rudimentary or deficient use of language • may have errors in grammar, usage, and mechanics that interfere with meaning
0
<ul style="list-style-type: none"> • response is totally incorrect or irrelevant

Grade 11 Writing Rubric – Reflective Essay

6
<ul style="list-style-type: none"> • purpose and context are engaging • intentionally organized, with a progression of ideas • analyzes a condition or situation using rich and insightful elaboration • distinctive voice, tone, and style enhance meaning • consistent application of the rules of grade-level grammar, usage, and mechanics
5
<ul style="list-style-type: none"> • purpose and context are clear • well organized and coherent throughout, with a progression of ideas • analyzes a condition or situation using meaningful details/elaboration • uses language effectively; uses a variety of sentence structures • consistent application of the rules of grade-level grammar, usage, and mechanics
4
<ul style="list-style-type: none"> • purpose and context are evident • generally organized and coherent • explains a condition or situation using relevant details • uses language adequately; uses correct sentence structures • may have some errors in grammar, usage, and mechanics
3
<ul style="list-style-type: none"> • writing has a general purpose • some sense of organization; may have lapses in coherence • addresses a condition or situation; some relevant details support purpose • uses language adequately; may show little variety of sentence structures • may have some errors in grammar, usage, and mechanics
2
<ul style="list-style-type: none"> • attempted or vague purpose • attempted organization; lapses in coherence • may state a condition or situation; generalized, listed, or undeveloped details • may lack sentence control or may use language poorly • may have errors in grammar, usage, and mechanics that interfere with meaning
1
<ul style="list-style-type: none"> • minimal evidence of purpose • little or no organization • few or random details • rudimentary or deficient use of language • may have errors in grammar, usage, and mechanics that interfere with meaning
0
<ul style="list-style-type: none"> • response is totally irrelevant

Grade 11 Writing Rubric –Report Writing

6
<ul style="list-style-type: none"> • purpose is clear throughout; strong focus/controlling idea OR strongly stated purpose focuses the writing • intentionally organized for effect • fully developed details, rich and/or insightful elaboration supports purpose • distinctive voice, tone, and style enhance meaning • consistent application of the rules of grade-level grammar, usage, and mechanics
5
<ul style="list-style-type: none"> • purpose is clear; focus/controlling idea is maintained throughout • well organized and coherent throughout • details are relevant and support purpose; details are sufficiently elaborated • strong command of sentence structure; uses language to enhance meaning • consistent application of the rules of grade-level grammar, usage, and mechanics
4
<ul style="list-style-type: none"> • purpose is evident; focus/controlling idea may not be maintained • generally organized and coherent • details are relevant and mostly support purpose • well-constructed sentences; uses language well • may have some errors in grammar, usage, and mechanics
3
<ul style="list-style-type: none"> • writing has a general purpose • some sense of organization; may have lapses in coherence • some relevant details support purpose • uses language adequately; may show little variety of sentence structures • may have some errors in grammar, usage, and mechanics
2
<ul style="list-style-type: none"> • attempted or vague purpose • attempted organization; lapses in coherence • generalized, listed, or undeveloped details • may lack sentence control or may use language poorly • may have errors in grammar, usage, and mechanics that interfere with meaning
1
<ul style="list-style-type: none"> • minimal evidence of purpose • little or no organization • random or minimal details • rudimentary or deficient use of language • may have errors in grammar, usage, and mechanics that interfere with meaning
0
<ul style="list-style-type: none"> • response is totally incorrect or irrelevant

Grade 11 Writing Rubric – Persuasive Writing

6
<ul style="list-style-type: none"> • purpose/position is clear throughout; strong focus/position; OR strongly stated purpose/opinion focuses the writing • intentionally organized for effect • fully developed arguments and reasons; rich, insightful elaboration supports purpose/opinion • distinctive voice, tone, and style effectively support position • consistent application of the rules of grade-level grammar, usage, and mechanics
5
<ul style="list-style-type: none"> • purpose/position is clear; stated focus/opinion maintained consistently throughout • well organized and coherent throughout • arguments/reasons are relevant and support purpose/opinion; arguments/reasons are sufficiently elaborated • strong command of sentence structure; uses language to support position • consistent application of the rules of grade-level grammar, usage, and mechanics
4
<ul style="list-style-type: none"> • purpose/position and focus are evident, but may not be maintained • generally well organized and coherent • arguments are appropriate and mostly support purpose/opinion • well-constructed sentences; uses language well • may contain some errors in grammar, usage, and mechanics
3
<ul style="list-style-type: none"> • purpose/position may be general • some sense of organization; may have lapses in coherence • some relevant details support purpose; arguments are thinly developed • generally correct sentence structure; uses language adequately • may contain some errors in grammar, usage, and mechanics
2
<ul style="list-style-type: none"> • attempted or vague purpose/position • attempted organization; lapses in coherence • generalized, listed, or undeveloped details/reasons • may lack sentence control or may use language poorly • may have errors in grammar, usage, and mechanics that interfere with meaning
1
<ul style="list-style-type: none"> • minimal evidence of purpose/position • little or no organization • random or minimal details • rudimentary or deficient use of language • may have errors in grammar, usage, and mechanics that interfere with meaning
0
<ul style="list-style-type: none"> • response is totally irrelevant

Grade 11 Writing Rubric –Procedure

Text features can serve as organizational devices or as details that enhance meaning.

6
<ul style="list-style-type: none"> • purpose and context are clear; strong focus/controlling idea maintained throughout • intentionally organized for effect • fully developed details and elaborated steps support purpose • distinctive voice, tone, and style enhance reader’s understanding • consistent application of the rules of grade-level grammar, usage, and mechanics
5
<ul style="list-style-type: none"> • purpose and context are clear; focus/controlling idea is maintained throughout • well organized and coherent throughout • details are relevant and support purpose; steps are sufficiently explained • precise word choice; sentence structure/phrasing is appropriate • consistent application of the rules of grade-level grammar, usage, and mechanics
4
<ul style="list-style-type: none"> • purpose and context are evident; • generally organized and coherent • details are relevant, clear, and mostly support purpose; steps are explained • specific word choice; sentence structure/phrasing is appropriate • may have some errors in grammar, usage, and mechanics
3
<ul style="list-style-type: none"> • purpose is general • some sense of organization; may have lapses in coherence • some relevant details support purpose; some steps are identified • may use nonspecific language; sentences/ phrases may be unclear • may have some errors in grammar, usage, and mechanics
2
<ul style="list-style-type: none"> • attempted or vague purpose • attempted organization; lapses in coherence • generalized, list or undeveloped details; may identify steps • may use language poorly; sentence/phrasing may cause confusion • may have errors in grammar, usage, and mechanics that interfere with meaning
1
<ul style="list-style-type: none"> • minimal evidence of purpose • little or no organization • random or minimal details • rudimentary or deficient use of language • may have errors in grammar, usage, and mechanics that interfere with meaning
0
<ul style="list-style-type: none"> • response is totally incorrect or irrelevant

APPENDIX H: NECAP STANDARD SETTING GRADE 11 RATING FORMS -WRITING ROUNDS 1 AND 2 -

NECAP WRITING GRADE 11

Rating Form

Common Prompt: *Working*

ID _____

	Round 1				Round 2			
	SBP	PP	P	PWD	SBP	PP	P	PWD
A								
B								
C								
D								
E								
F								
G								
H								
I								
J								
K								
L								
M								
N								
O								
P								

APPENDIX I NECAP STANDARD SETTING GRADE 11 RATING FORM -WRITING FINAL ROUND -

NECAP Writing Grade 11 Final Round

ID _____

Starting Cutpoints

<p>Substantially Below Proficient Score Points</p> <p>First Last 2 3</p>	<p>Partially Proficient Score Points</p> <p>First Last 4 6</p>	<p>Proficient Score Points</p> <p>First Last 7 9</p>	<p>Proficient with Distinction Score Points</p> <p>First Last 10 12</p>
---	---	---	--

Final Round

Directions: Please enter the range of score points that fall into each achievement level category according to where you believe the cutpoints should be placed.

<p>Substantially Below Proficient Score Points</p> <p>First Last 2 ____</p>	<p>Partially Proficient Score Points</p> <p>First Last ____ ____</p>	<p>Proficient Score Points</p> <p>First Last ____ ____</p>	<p>Proficient with Distinction Score Points</p> <p>First Last ____ 12</p>
--	---	---	--

APPENDIX J: NECAP STANDARD SETTING EVALUATION SUMMARIES

NECAP Grade 11 Standard Setting—Mathematics 17 Evaluations completed

January 9-10, 2008

Sheraton Harborside Portsmouth Hotel, Portsmouth, NH

1. How would you rate the training you received?

<i>Rating Scale</i>	<i>Appropriate</i>	<i>Somewhat Appropriate</i>	<i>Not Appropriate</i>
Tally	17	0	0

2. How clear were you with the achievement level definitions?

<i>Rating Scale</i>	<i>Very Clear</i>	<i>Clear</i>	<i>Somewhat Clear</i>	<i>Not Clear</i>
Tally	8	7	2	0

3. How do you feel about the length of time of this meeting for setting achievement standards?

<i>Rating Scale</i>	<i>About Right</i>	<i>Too little time</i>	<i>Too much time</i>
Tally	17	0	0

Comment: Try to do more on day 1

4. What factors influenced the standards you set?

(Circle the most appropriate rating from 1 = Not at all Influential to 5 = Very Influential)

Rating Scale:	1	2	3	4	5
The achievement level definitions			3	8	6
The assessment items			1	5	11
Other panelists Comment: Excellent discussions!		1	7	8	1
Impact Data	4	2	7	3	1
My experience in the field			1	4	12
Other (<i>please explain</i>)					

or each statement below, please circle the rating that best represents your judgment.

5. The training was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally				4	13

6. The achievement level definitions were:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally		1	5	5	6

For each statement below, please circle the rating that best represents your judgment.

7. Reviewing the assessment materials was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally			1	2	14

8. The discussion with other panelists was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally			1	6	10

9. The standard setting task was:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally			2	7	8
Comment: Clearer as time went on.					

10. My level of confidence in setting cut-points is:

Rating scale	1 <i>Very Low</i>	2	3	4	5 <i>Very High</i>
Tally			1	10	5

*one panelist did not answer

11. How could the standard setting process have been improved?

- Some discussions were too long and off-track
- I calibrated with my special-ed chair before coming by looking at student work. (I wanted her perspective, in my mind's eye).
- Change the standards for special-ed. That's a big issue that affects the scores quite a bit.
- Everything was superb.
- If I did it again, it would be much easier.
- Maybe a "mock training" first with pilot data or grade 8 data
- More open process that allows for participants to spend more time with different aspects of the program, (allow more time for discussion or item rating, as needed by individual)
- I'm not sure the process could be improved. The facilitators were fabulous. Everyone was focused on leading us through with a thorough understanding.
- Validation with another, similar group of teachers.
- I would start by setting the cut points for Proficient, then worry about Substantially Below to Partially Proficient later.
- Quickly review the GSEs
- Whipping around to hear quick opinions on different topics from *all* members (some chose not to share)
- A bit more time to talk.
- Reviewing the GSEs before taking the test and setting standards.

12. Please use the space below to provide any additional comments or suggestions about the standard setting process

Achievement Level Definitions

- Achievement Level Definitions need measurable benchmarks/goals. They are too subjective in their present form.
- It seemed (very much) as though our feedback on Achievement Level Definitions was not well received or sincerely wanted.
- The Achievement Level Definitions [discussion] should have been typed up and put back to the group before the first run through cut points because our definitions of the 4 “buckets” were not totally clear.

Facilitators/Presenters

- Great group, Phil was fun & a good facilitator.
- Phil was a *great* facilitator, very helpful, informative, and knowledgeable.
- A *very* helpful conference to attend. Interaction with colleagues from my state and other two states is invaluable.
- Phil was terrific in maintaining climate & focus.
- Excellent task facilitator and great moderators from each state.
- People from Measured Progress were extremely professional and informed. I have almost unlimited positive things to say about the people from Measured Progress, they were all awesome.

Other

- Organization of the activities was good.
- Small group & whole group mixtures may be rich.
- A *very* helpful conference to attend.
- Interaction with colleagues from my state and other two states is invaluable.
- I would encourage more people to become involved in the process.
- Excellent accommodations.
- Good test, besides the special-ed issue, I think things are going in the right direction. We do need teachers that are qualified and do their job correctly.

NECAP Grade 11 Standard Setting—Reading - 15 Evaluations completed

January 9-10, 2008

Sheraton Harborside Portsmouth Hotel, Portsmouth, NH

1. How would you rate the training you received?

<i>Rating Scale</i>	<i>Appropriate</i>	<i>Somewhat Appropriate</i>	<i>Not Appropriate</i>
Tally	15		

2. How clear were you with the achievement level definitions?

<i>Rating Scale</i>	<i>Very Clear</i>	<i>Clear</i>	<i>Somewhat Clear</i>	<i>Not Clear</i>
Tally	5	8	2	

3. How do you feel about the length of time of this meeting for setting achievement standards?

<i>Rating Scale</i>	<i>About Right</i>	<i>Too little time</i>	<i>Too much time</i>
Tally	14	1	

4. What factors influenced the standards you set?

(Circle the most appropriate rating from 1 = Not at all Influential to 5 = Very Influential)

Rating Scale:	1	2	3	4	5
The achievement level definitions		1	3	4	7
The assessment items				8	7
Other panelists			3	10	2
Impact Data		1	6	8	
My experience in the field				6	9
Other (please explain)					

For each statement below, please circle the rating that best represents your judgment.

5. The training was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally				6	9

6. The achievement level definitions were:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally			2	8	5

Comment: rated a 4 only after our elaboration of the definitions

For each statement below, please circle the rating that best represents your judgment.

7. Reviewing the assessment materials was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally				1	14

8. The discussion with other panelists was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally			1	4	10

9. The standard setting task was:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally			2	9	4

10. My level of confidence in setting cut-points is:

Rating scale	1 <i>Very Low</i>	2	3	4	5 <i>Very High</i>
Tally			1	12	2

11. How could the standard setting process have been improved?

- Perhaps matching the items with our elaboration of the Achievement Level Definitions. (Although, I understand that this might have been too time consuming).
- I'd like to see the data while making my bookmarks.
- I was unclear about the big picture and what was expected of me.
- I think a clearer outline of the tasks would have been helpful prior to beginning the process. I was very confused about how the bookmark process worked, however it did become clear over time.
- People need more information up front, before attending and at the beginning of the process, the big picture.
- Up front outline of entire process, maybe a rough outline of time frame to keep discussion focused.
- Fewer tangential conversations, discussions, and arguments which were not relevant to the task.
- Fewer people coming in and out of the room.
- Process went very smoothly. David was very diplomatic as a facilitator and did an excellent job clarifying what was sometimes a gray area.
- Facilitators needed to respect the situation and not talk (even in whispers) during the time we were taking the test.
- Reduce noise from adjacent room (3 comments)

12. Please use the space below to provide any additional comments or suggestions about the standard setting process

Process

- Keep one test the whole time
- There should be an outlet for discussion of test items. Very frustrating and unnatural to push it out of discussion.
- I would have liked to have seen an item-by-item analysis of students' performance on all 52 questions (percentages of right/wrong for MC items or 1-4 score percentages on CR items)
- It would be interesting to see how we would rank the skill difficulty of the items with how the students actually performed.

Involvement in standard setting

- I was very happy to have participated in the standard setting process.
- Very interesting professionally.
- This process was extremely helpful to me as an educator. One crucial piece of information that I now have is the way that data is used to assess my students.
- Overall I really enjoyed the process and am looking forward to hearing the results. I appreciate that classroom teachers are so actively involved.

Other

- David did an excellent job leading us through the process.
- Our instructor needed to stop discussions that were getting off task. Too much time spent arguing the test itself.
- Accommodations were lovely
- The venue and support were outstanding.
- Less noisy setting.
- Closer proximity of participants when discussing
- Less interruptions during meetings (fewer people entering /leaving, i.e. observers)
- People were friendly and the professionals extremely knowledgeable.

NECAP Grade 11 Standard Setting –Writing - 16 Evaluations completed

January 9-10, 2008

Sheraton Harborside Portsmouth Hotel, Portsmouth, NH

1. How would you rate the training you received?

<i>Rating Scale</i>	<i>Appropriate</i>	<i>Somewhat Appropriate</i>	<i>Not Appropriate</i>
Tally	15	1	

2. How do you feel about the length of time of this meeting for setting achievement standards?

<i>Rating Scale</i>	<i>About Right</i>	<i>Too little time</i>	<i>Too much time</i>
Tally	14		2

Comments: Time within the 2 days could have been better structured. (rated *About Right*)
Day 2 process should be set up so there is less waiting around. (rated *Too much time*)

3. What factors influenced the standards you set?

(Circle the most appropriate rating from 1 = Not at all Influential to 5 = Very Influential)

Rating Scale:	1	2	3	4	5
The achievement level definitions				7	9
The writing prompts *one panelist did not answer	1		5	7	2
The student responses			1	6	9
Other panelists	3	1	8	4	
My experience in the field				6	10
Other (<i>please explain</i>)			1		1
Explanations of <i>Other</i> ratings: 3 = The rubrics 5 = Grade 11 writing rubric, because it is the standard for what is best and what is not acceptable. Helped me gauge writing quality. Other comment (with no rating): the group process					

For each statement below, please circle the rating that best represents your judgment.

4. The training was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally				6	10

5. The scoring rubrics were:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally			1	3	12

For each statement below, please circle the rating that best represents your judgment.

6. Reviewing the assessment materials was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally				6	10

7. The discussion with other panelists was:

Rating scale	1 <i>Not Useful</i>	2	3	4	5 <i>Very Useful</i>
Tally			2	5	9

8. The standard setting task was:

Rating scale	1 <i>Not Clear</i>	2	3	4	5 <i>Very Clear</i>
Tally			2	5	9
Comment: became more so as we progressed (rated 4)					

9. My level of confidence in setting cut-points is:

Rating scale	1 <i>Very Low</i>	2	3	4	5 <i>Very High</i>
Tally			3	11	2

10. How could the standard setting process have been improved?

Process

- Do not give the scores or rank the papers, it really influenced the work. (5 comments)
- Downtime, waiting between matrix pieces feels wasted. (3 comments)
- Consider not giving in advance what the ELA folks had set for their cut points. I feel like it may have influenced the group.
- By doing 4 essays on day 2 I really lost steam.
- It was clearly explained, executed efficiently, and gave us the sense that what we think matters. In short, little needs improvement.
- I felt the time allotted for each task was sufficient.
- Spacing/enough time to do task and more breaks to rest/refresh.

Scoring

- It seems that discrepancies in scores could have been discussed further to glean the reasons between the various scores. Allow more discussion in this area.
- I would spend more time on authentic scoring rather than providing the scores while we read the pieces. It biases the reader.

Other

- I felt Measured Progress really tried to hear and answer our concerns.
- Talk at meals helped make me feel comfortable with others so I could share my opinions in sessions.
- Some discussions went on too long
- Would have like to have the schedule before coming
- Impressed with the whole process. The info and work done prior to our meeting was well thought out and professional and grounded in the right values to challenge and support good student learning and promote the process we came here for.
- The very loud fan made it difficult to hear comments from some folks across the room.

11. Please use the space below to provide any additional comments or suggestions about the standard setting process

Cut points and Process

- I am still uncomfortable with 7 as a cut score for proficient. I would be more comfortable with a definition corresponding to a score point—even if it were 3.5
- While I understand that ours is not the only factor in determining the cut points, the facilitator's comments about the process have led me to feel that we are merely validating points that have already been at least mentally established by Measured Progress. I've not been pressured to change my rating, but I've been given the impression that I also won't influence the decision that will be made (at least not much).
- Some concern over the Partially Proficient and Proficient cut. If this were *only* about improving students skills, then it would be fine. However, we all know the repercussions of poor test scores. Perhaps erring on the side of generosity would help the big picture and not really hurt the students' progress.
- Do this *before* papers are scored.
- Provide work samples/prompts for participants to bring back to their schools/districts

Overall concerns

- My concern goes back to the scoring that was already completed before we came to this task. Just with the number of pieces we saw, several of us saw papers "on the cusp". In some cases, that 6/7 rated piece could be making/breaking a school district. With the stakes so high, I'm still not completely comfortable with scoring. I appreciate Linda's [NH DOE] explanation, and I trust her, but I'm still uneasy.
- I hope the DOE folks from the three states will hear our concerns about what may influence the way students prepare for and see the test. Teachers' language in the classroom (prompt vs task, essay vs response) will be lenses for how students tackle assignments. The language of John Collins, Nancy Atwell, Peter Elbow (sp?) impacts how teachers teach and students respond.
- We have heard repeatedly that NECAP is a snapshot, and no where is this more applicable than in the writing portion. Continue to stress this point in all discussions of results with state and local education officials. It's imperative that parents and press understand this fact.
- Work on the word "prompt"

Facilitators and facility

- The facilitators were excellent and I enjoyed working with them.
- Tim did a fantastic job of leading us all through quite a complicated process.
- I applaud the work of the representatives from the three states.
- MP staff was friendly and very helpful.
- The facility was outstanding.

Other

- This was a valuable experience for me and I enjoyed it.
- Thank you for believing that classroom teachers actually have something valuable to contribute to the process.
- Interesting, helped me understand the NECAP testing process and scoring. I will be armed with this new found knowledge when I assess my classroom instruction.
- I have never participated in an exercise quite like this, but I feel as though the process was valid.
- Have more SPED folks involved.
- Overall this was a useful experience.
- I attended standard setting for grade 8 and had a terrible go of it. This was *much better*. We had clear directions and a clear task. I really enjoyed the process.

APPENDIX K: NECAP STANDARD SETTING PANELISTS

New Hampshire

Reading

First Name	Last Name	School/Association Affiliation	Position
Susan	Dean-Olsen	Kingswood Regional High School	English Language Arts Teacher/Coordinator
Jack	Finley	Franklin High School	English Language Arts Teacher
Joanne	O'Connor	Pinkerton Academy	Special Education Teacher
Jeanne	Provender	Nashua (retired)	English Language Arts Teacher
Chris	Saunders	Nashua High School	English Language Arts Teacher
Michael	Williamson	Hollis/Brookline High School	English Language Arts Teacher

Mathematics

First Name	Last Name	School/Association Affiliation	Position
Linda	Belmonte	Bedford High School	Dean
Tracy	Bricchi	Kearsarge School District	Mathematics Coordinator
Marina	Capen	Souhegan High School	Mathematics Teacher
Robert	Comey	Memorial High School	Mathematics Teacher
Matt	Cygan	Memorial High School	Mathematics Teacher
Rob	Lukasiak	Independent Consultant	Mathematics Consultant for CEIL

Writing

First Name	Last Name	School/Association Affiliation	Position
Carrie	Costello	Conway High School	English Language Arts Teacher
Kim	Lindley-Soucy	Londonderry High School	English Language Arts Curriculum Coordinator
Meg	Petersen	Plymouth University	Plymouth Writing Project
Jean	Shankle	Milford High School	English Language Arts Teacher
Ruth Ellen	Vaughn	Farmington	English Language Arts Curriculum Coordinator
Ann	West	Pinkerton Academy	English Department Chair

Rhode Island

Reading

First Name	Last Name	School/Association Affiliation	Position
Patricia	Armstrong	East Providence High School	Department Chair
Jill	Burke	Chariho High School	English Language Arts Teacher
Jean	Dietrich	Community College of Rhode Island	English Language Arts Teacher
Rebecca	Moore	Mt. Hope High School	English Language Arts Teacher
Sharon	Solway	Mt. Hope High School	English Language Learner Teacher

Mathematics

First Name	Last Name	School/Association Affiliation	Position
Michelle	Brousseau-Cavallaro	East Providence High School	Department Chair
Linda	Curtin	Hope Arts High School	Mathematics Teacher
Jean	Mollicone	Mt. Hope High School	Department Chair
Suzanne	Ross Walker	Woonsocket High School	AP Calculus Teacher
Monique	Rousselle-Condon	West Warwick High School	Mathematics Teacher

Writing

First Name	Last Name	School/Association Affiliation	Position
David	Groccia	North Providence High School	English Language Arts Teacher
Emmanuel	Vincent	E-Cubed Academy	Special Education Teacher
Jeff	Miner	Toll Gate High School	Department Chair
David	Schofield	Lincoln Senior High School	Department Chair

Vermont

Reading

First Name	Last Name	School/Association Affiliation	Position
Alan	Crowley	Missisquoi Valley Union	English Language Arts Teacher and Department Leader
Sue	Boardman	Brattleboro Union High School	English Language Arts Teacher
Colleen	Fiore	Long Trail School	English Language Arts Teacher and Special Services Director
Sandy	Frizzell	North Country Union High School	English Language Arts Teacher
Katie	Lenox	Colchester High School	English Language Arts Teacher
Marilyn	Woodard	Mt. Anthony Union High School	English Language Arts Teacher and Department Chair

Mathematics

First Name	Last Name	School/Association Affiliation	Position
Laurie	Camelio	Mt. Anthony Union High School	Mathematics Teacher and Department Chair
Mike	Caraco	Burr and Burton Academy	Mathematics Teacher and Department Chair
Nancy	Disenhaus	U-32 High School	English Language Arts Teacher
Sharon	Fadden	Danville High School	Mathematics Teacher
Erik	Jacobson	Windham Northeast Supervisory Union	Mathematics Teacher
John	Pandolfo	Spaulding High School	Mathematics Teacher & Department Head

Writing

First Name	Last Name	School/Association Affiliation	Position
Teri	Appel	Brattleboro Union High School	English Language Arts Teacher and Literacy Network Leader
Renee	Berthiaume	North Country Union High School	Literacy Coach & Language Arts Department Liaison
Erin	McGuire	Colchester High School	English Humanities Teacher
Peter	Riegelman	St. Albans	English Language Arts Teacher
Susan	Soltau	Essex High School	Mathematics Teacher & Co-chair Mathematics Department

